Technisches Whitepaper

HPE GreenLake for File Storage Architektur

Inhalt

Kurzübersicht 4
Zielgruppe 4
Hardwarearchitektur 4
Architektur der Computing-Nodes 5
DBox-Gehäuse (für SCM und NVMe) 5
Front-End-Konnektivität 6
Softwarearchitektur 6
DASE-Architektur 6
Serverorientiertes Softwaredesign 7
Dateisystem, Protokolle und protokollübergreifende Funktionen 8
Globaler Namespace 10
Leistung 10
Datenpfad 11
Controller-Pooling, Servicequalität und Mandantenfähigkeit 12
Höhere NFS-Leistung 13
Datenspeicherkapazität 14
Datenreduzierungstechnologien 14
Defragmentierung 15
Kontingente 15
Datenresilienz und Zuverlässigkeit 16
Schreib-Datenpfad 16
Schutz durch lokal decodierbares Erasure Coding 16
Hardware-Fehlertoleranz 16
Online-Upgrades 17
Snapshots, Aufbewahrung und Unveränderlichkeit 18
Replikation 18
Sicherheit 21
Datenberechtigung und -authentifizierung 21
Verschlüsselung ruhender Daten 22
Auditierung 22
Mandantenfähigkeit 22
Sicheres Cloud-basiertes Management 22
Sonstige Sicherheitsfunktionen 22
Speichermanagement 23
Data Services Cloud Console 23
Onboard UI 24

HPE iLO 24
API 24
Unterstützung von Container-Workloads 24
Zusammenfassung 24
Anhang 25

Kurzübersicht

Die Hybrid Cloud mit der zunehmenden Verbreitung künstlicher Intelligenz in Unternehmen stellt uns vor mehrere Herausforderungen: eine explosionsartige Zunahme unstrukturierter Daten und die allgegenwärtige und ständig steigende Nachfrage nach der Erschließung von Geschäftswert bei gleichzeitiger zukunftssicherer Gestaltung der IT-Infrastruktur mit Scale-Out-Architekturen, die wirklich auf die Größe von KI skalierbar sind.

Hewlett Packard Enterprise begegnet diesen Herausforderungen mit der Lösung **HPE GreenLake for File Storage**, die auf der disaggregierten **HPE Alletra Storage MP** Plattform in Varianten mit Standard- und hoher Dichte beruht, gekoppelt mit dem Betriebssystem HPE GreenLake for File Storage. Diese Lösung basiert auf einer einzigartigen DASE-Architektur (Disaggregated and Shared Everything), um Unternehmensleistung im KI-Maßstab zu liefern und gleichzeitig auf intuitive Weise das Beste der Cloud mit einer der effizientesten Kapazitäten und Leistungen pro Rack-Einheit bereitzustellen.

Um die KI-Leistung im KI-Maßstab weiter zu validieren, hat HPE GreenLake for File Storage Folgendes erreicht: NVIDIA DGX BasePOD Zertifizierung und NVIDIA OVX Speichervalidierung zusätzlich zu NVIDIA GPUDirect Storage (GDS) sowohl in Varianten mit Standard- als auch hoher Dichte.

Zielgruppe

Dieses Handbuch richtet sich an Lösungsarchitekten, Außendienstberater, IT-Spezialisten und alle anderen, die diese Informationen bei der Entwicklung und Implementierung einer HPE GreenLake for File Storage Lösung benötigen könnten. Das Dokument bietet eine Übersicht über die Lösung mit einem detaillierten technischen Einblick in die Hardware- und Softwarearchitektur. Von den Lesern wird erwartet, dass sie die Grundlagen der Dateispeicherung, der x86-Architektur, des grundlegenden Networkings, der Authentifizierungsmethoden und von Protokollen wie NFS und SMB verstehen.

Hardwarearchitektur

HPE Alletra Storage MP hat drei zugrunde liegende Komponenten.

1. Computing-Gehäuse, auch bezeichnet als **CBox**
2. Storage-Gehäuse oder JBOF, auch bezeichnet als **DBox**
3. Back-End-NVMe über Ethernet-Fabric

HPE Alletra Storage MP gibt es als Varianten mit **Standarddichte** und **hoher Dichte**. Die zugrunde liegende Speicherarchitektur ist von Natur aus disaggregiert und bietet die Flexibilität, Computing und Storage je nach Bedarf unabhängig voneinander zu skalieren, unabhängig von den verwendeten zugrunde liegenden Modulen mit Standard- oder hoher Dichte. Sowohl die HPE Alletra Storage MP-Module mit hoher Dichte als auch mit Standarddichte wurden mit NVIDIA DGX Systemen zertifiziert und mit NVIDIA OVX Systemen validiert.

Module mit Standarddichte sind mit der CBox in einem 2U-Gehäuse in Standard-Racktiefe mit 2 CNodes innerhalb der Cbox erhältlich. Für Datenspeicher sind die Module mit Standarddichte in einem 2U-DBox-Gehäuse erhältlich, ebenfalls in Standard-Racktiefe, und mit 2 DNodes innerhalb der DBox. Das Back-End-NVMe-over-Ethernet-Fabric ist als ein redundantes HPE Aruba Networking Switch-Paar mit 32 Ports und 100 GbE erhältlich und es kann durch die Option zur Erweiterung des Back-End-Fabrics mit Spine und Leafs auf einen größeren Footprint erweitert werden.

Module mit hoher Dichte sind mit der CBox in einem 2U-Gehäuse in Standardtiefe erhältlich, wobei 4 CNodes innerhalb der CBox den Formfaktor sowie Southbound- und Northbound-Verbindungen verdoppeln. Zur Speicherung werden die Module mit hoher Dichte in einem 1U-DBox-Gehäuse in einem Rack mit erweiterter Tiefe mit 4 DPUs geliefert, die innerhalb der DBox auch als Smart NICs bezeichnet werden, wodurch sich auch der Formfaktor und die Verbindungen zum Back-End-Fabric verdoppeln. Das Back-End-NVMe-over-Ethernet-Fabric ist als redundantes NVIDIA® Spectrum SN4600 Switch-Paar mit 64 Ports und 100 GbE erhältlich. Es kann

mit der Option zur Erweiterung des Back-End-Fabrics mit Spine und Leafs auf einen größeren Footprint erweitert werden.

Tabelle 1. Vergleich der zugrunde liegenden Bausteine

Plattform-Komponente	Eigenschaften	HPE Alletra Storage MP mit Standarddichte	HPE Alletra Storage MP mit hoher Dichte
Computing-Gehäuse	Gehäuse	2U-Gehäuse mit 2 Nodes. Chassis mit Standardtiefe	2U-Gehäuse mit 4 Nodes. Chassis mit Standardtiefe
	Node	1–32C Zentraleinheit (AMD) 256 GB RAM, 2 PCIe4-Karten mit dualen 100-G-Ports	2–16C Zentraleinheit (Intel®) 256 GB RAM, 2 PCIe4-Karten mit dualen 100-G-Ports
Storage-Gehäuse	Gehäuse	2U-Gehäuse mit 2 Nodes. Chassis mit Standardtiefe	1U-Gehäuse mit 2 Nodes. Chassis mit Standardtiefe
	Node	1 PCIe4-Karte mit dualem 100-GbE-Port (4 Ports pro DBox)	1 PCIe4-Karte mit dualem 100-GbE-Port (8 Ports pro DBox)
	Kapazitäts-Optionen	154-TB-(Raw)-DBox mit 8-TB-Laufwerken, 308-TB-(Raw)-DBox mit 15-TB-Laufwerken & 550-TB-(Raw)-DBox mit 30-TB-Laufwerken	338-TB-(Raw)-DBox mit 15-TB-Laufwerken & 1352-TB-(Raw)-DBox mit 60-TB-Laufwerken
Back-End-NVMe-oF n/a	n/a	1U HPE Aruba Networking CX 8325 Switch-Paar mit 100 GbE und 32 Ports	2U NVIDIA Spectrum SN4600 Switch-Paar mit 100 GbE und 64 Ports

Architektur der Computing-Nodes

HPE GreenLake for File Storage, basierend auf HPE Alletra Storage MP Computing-Nodes, ist ein wichtiger Teil des Systems. Ein einzelnes CBox Computing-Gehäuse besteht aus zwei Computing-Nodes in der Variante mit Standarddichte und 4 CNodes in der Variante mit hoher Dichte. Von jedem CNode aus sind zwei 100-GbE-Ports mit dem Back-End-NVMe-over-Ethernet-Fabric verbunden, und die anderen beiden 100-GbE-Anschlüsse sind mit dem Ethernet- oder InfiniBand-Netzwerk des Kunden verbunden. Für die Client-Konnektivität werden zusätzlich zur 100-GbE-Konnektivität auch 40-GbE-, 25-GbE- und 10-GbE-Konnektivität durch entsprechende Transceiver mit OM4-LC-Kabeln oder entsprechenden aktiven optischen Kabeln unterstützt. Weitere Informationen entnehmen Sie bitte den QuickSpecs.

DBox-Gehäuse (für SCM und NVMe)

In HPE GreenLake for File Storage, das auf HPE Alletra Storage MP basiert, besteht das JBOF DBox-Gehäuse aus NVMe-SSD-Laufwerken und Storage-Class-Memory-(SCM)-Laufwerken.

HPE Alletra Storage MP in der DBox mit **Standarddichte** gibt es mit 24 SFF-Laufwerkssteckplätzen. Etwa 2 % der Gesamtkapazität jeder DBox sind für SCM reserviert, der wie NVRAM verwendet wird. Dieser verteilte SCM-Layer bildet auch den Metadatenspeicher. In einer DBox mit 7,68-TB- oder 15,36-TB-Laufwerken sind 4 Laufwerkssteckplätze für SCM-Laufwerke reserviert. Die restlichen 20 Steckplätze sind für NVMe-Laufwerke vorgesehen. In einer DBox mit 30-TB-Laufwerken sind 6 Laufwerkssteckplätze für SCMs reserviert, die restlichen 18 Steckplätze für NVMe-Laufwerke. Jeder DNode besteht aus einem Host-Bus-Adapter mit zwei Ports und 100 GbE (R7C84A). Zwei 100-GbE-Ports sind mit den HPE Aruba Networking Back-End-Switches verbunden. So sind in jeder DBox vier 100-GbE-Ports mit den Back-End-Switches verbunden. Details zu den Systembausteinen finden Sie in den QuickSpecs.

Die HPE Alletra Storage MP in DBox mit **hoher Dichte** verfügt über 22 linealbasierte NVMe-SSD-Laufwerke und 8 SCM-Laufwerke. Etwa 2 % der Gesamtkapazität jeder DBox sind für SCM reserviert, der wie NVRAM verwendet wird. Dieser verteilte SCM-Layer bildet auch den Metadatenspeicher. Anders als bei der Standarddichte werden

die NVMe-SSDs und SCM-Laufwerke unabhängig von der Laufwerkskapazität reserviert. Die DBox mit 15-TB-NVMe-Laufwerken verfügt über 1,6-TB-SCMs, und in einer DBox mit 60 TB NVMe haben SSD-Laufwerke 3,2-TB-SCM-Laufwerke. Jede DPU in der DBox mit hoher Dichte besteht aus einem 100-GbE-Hostbusadapter mit zwei Ports. Zwei 100-GbE-Ports sind mit den Back-End-Switches NVIDIA Spectrum SN4600 verbunden. So sind in jeder DBox acht 100-GbE-Ports mit den Back-End-Switches verbunden. Details zu den Systembausteinen finden Sie in den QuickSpecs.

Front-End-Konnektivität

HPE GreenLake for File Storage unterstützt Ethernet und InfiniBand für Front-End-Konnektivität auf HPE Alletra MP Plattformvarianten mit Standard- als auch hoher Dichte. HPE GreenLake for File Storage kann mit ausschließlich Ethernet-Optionen auf allen CBoxes oder mit ausschließlich InfiniBand auf allen CBoxes konfiguriert werden. Alternativ kann es im gemischten Modus mit Ethernet und InfiniBand konfiguriert werden. Die Auswahl erfolgt auf CBox-Ebene.

HPE GreenLake for File Storage, das auf Modulen mit **Standarddichte** basiert, gibt es mit zwei Optionen für die Front-End-Konnektivität. Für Lösungen mit ausschließlich Ethernet-Konfigurationen sollten R7C84A-Adapter bestellt werden, während für Lösungen mit ausschließlich InfiniBand und einer Mischung aus InfiniBand und Ethernet S2H33A-Adapter bestellt werden sollten. HPE GreenLake for File Storage, das auf Modulen mit **hoher Dichte** basiert, verfügt hingegen über Ethernet/InfiniBand-Dual-Purpose-Adapter.

Im Wesentlichen dienen die über OCA verfügbaren Optionen zur Aktivierung der Standarddichte und die Wahl zwischen Ethernet und InfiniBand. Die Variante für Cboxen mit hoher Dichte wird standardmäßig mit InfiniBand/Ethernet-fähigen Adaptern ausgeliefert. Diese bieten die Wahl zwischen InfiniBand und Ethernet-Konnektivität zum Installationszeitpunkt.

Softwarearchitektur

Die HPE GreenLake for File Storage Lösung unterscheidet sich durch eine Reihe von Merkmalen von anderen Scale-out-NAS-Angeboten. Dank der einzigartigen DASE-Software-Architektur können Sie klein anfangen und mit unabhängiger Kapazitäts- und Leistungsskalierung einen Maßstab erreichen, der mit Shared-Nothing-Systemen unmöglich war. Diese Lösung bietet branchenweit ungeschlagene Datenreduktionstechnologien mit einer Reihe von einzigartigen Techniken.

DASE-Architektur

Das DASE-Architekturkonzept ist ein großer Sprung in der Welt der verteilten Systeme. Diese Shared-Everything-Architektur bietet revolutionäre Vorteile. Im nachfolgenden Abschnitt werden einige der wichtigsten Alleinstellungsmerkmale beschrieben.

Für lineare Skalierung entwickelt

Eine Shared-Nothing-Architektur stützt sich in der Regel stark auf die Möglichkeit, jeden Controller-Node heraufzuskalieren, um eine hohe Leistung zu erzielen. Das bedeutet, dass die Verantwortung für die Daten auf die Controller-Nodes aufgeteilt wird. Diese Art der Architektur mag in einem kleineren System sinnvoll gewesen sein, um hohe Leistung zu garantieren, doch im Zeitalter der **Datenexplosion** werden immer mehr Unzulänglichkeiten von Shared-Nothing-Architekturen durch Scale-out-Systeme offenbart. So erfordert eine Shared-Nothing-Architektur insbesondere einen Ost-West-Datenverkehr, der einen Overhead darstellt und mit zunehmender Menge von Controller-Nodes schnell in die Höhe schießt. Durch eine DASE-Architektur wird zur Behebung dieser Unzulänglichkeiten beigetragen. Ohne Ost-West-Datenverkehr entsteht beim Heraufskalieren kein Latenz-Overhead. Sie können unabhängig CBoxes hinzufügen, um mehr Leistung zu erzielen, und Sie können ebenso unabhängig DBoxes hinzufügen, ohne die Rechenleistung erweitern zu müssen, da die Verantwortung für den Datenspeicher bei dieser Lösung nicht länger an Controller-Nodes oder Node-Paare gebunden ist.

Kein Ost-West-Datenverkehr

Bei zu vielen gemeinsamen Status der Controller-Nodes wird eine Shared-Nothing-Architektur möglicherweise durch Ressourcenkonflikte zwischen den Controller-Nodes beeinträchtigt. Das bedeutet eine erhebliche Menge

an Crosstalk zwischen Servern. Ein Nebeneffekt besteht darin, dass Wiederherstellungen von Servern deutlich mehr Zeit in Anspruch nehmen und so die Gesamtleistung und Zuverlässigkeit des Systems beeinträchtigen können. Bei einer DASE-Architektur kommt es gar nicht erst zu diesen Konflikten, da alle CNodes die gleichen Pfade zur gesamten Datenstruktur nutzen. Alle CNodes in einem Cluster können die I/Os auf dieselbe Weise betreiben und bedienen. So lassen sich I/Os gleichmäßig auf alle CNodes verteilen, ohne Konflikte zu verursachen. Das Resultat: Bei dieser Architektur ist aufgrund des Ausfalls eines Computing-Nodes keine Wiederherstellung erforderlich.

Flüchtiges Computing

Bei der HPE GreenLake for File Storage Lösung sind die CNodes statusfrei, weil die DASE-Architektur eine vollständige Unabhängigkeit des Computing-Layers von Metadaten ermöglicht. Metadaten werden immer in einem persistenten verteilten SCM-Layer (auch NVRAM genannt) in den DBoxes gespeichert. Alle CNodes im Datei-Cluster können jedes Laufwerk im Cluster so sehen, als wären sie direkt angeschlossen. Außerdem haben alle CNodes Zugriff auf den verteilten SCM-Layer, auf dem die Metadaten gespeichert werden. So entsteht kein Crosstalk zwischen den CNodes, es sind keine Serverwiederherstellungen nach einem Ausfall notwendig usw. Der Abschluss eines I/O-Vorgangs trägt dazu bei, die Festschreibung der Daten auf dem persistenten Datenspeicher (SCM) zu garantieren. Dieses Konzept erleichtert die Cache- und Arbeitsspeicherbereinigung bei einem Komponentenausfall und verhindert so potenzielle Datenverluste oder Inkonsistenzen in der Datenstruktur. Eine Reihe von Mechanismen wie beispielsweise Atomic-Aktualisierungen, die Änderung von Baumstrukturen von unten nach oben, Sperren und Timeouts tragen dazu bei, die Konsistenz der Datenstruktur sicherzustellen, und der Zugriff auf die Datenstruktur wird von den CNodes gemeinsam genutzt.

In Tabelle 2 werden die wichtigsten Vergleichspunkte zusammengefasst.

Tabelle 2. Eine Zusammenfassung der Vergleiche zwischen Shared-Everything-Architekturen und Shared-Nothing-Architekturen

Herausforderungen bei Shared-Nothing-Architekturen	Merkmale von Shared-Everything-Architekturen	Vorteile von Shared-Everything-Architekturen
Beeinträchtigung der Latenz durch zunehmenden Overhead auf Scale-out-Systemen.	Es gibt keinen Ost-West-Datenverkehr zwischen CNodes.	Durch die Skalierung von CNodes entsteht kein Leistungs-Overhead. Die Latenz wird nicht beeinträchtigt, und der Durchsatz wird im Einklang mit dem System linear skaliert.
Es kommt zu Konflikten um Ressourcen zwischen CNodes.	CNodes haben einen gemeinsamen Zugriff auf die gesamte Datenstruktur.	Es kommt nur noch zu minimalen Konflikten unter CNodes.
Persistente Status (ständige Aufrechterhaltung der Cache-Kohärenz) ist schwer zu erzielen.	CNodes haben keinen persistenten Status, sehen immer denselben globalen Status und müssen keinen Cache aufrechterhalten.	Beim Ausfall eines CNodes kommt es zu keinen Datenverlusten oder Inkonsistenzen der Datenstruktur, und Ausfälle von Computing-Nodes sind deutlich leichter zu handhaben.

Serverorientiertes Softwaredesign

Der Software-Stack von HPE GreenLake for File Storage lässt sich vereinfacht in drei Layern darstellen:

– Firmware

– BETRIEBSSYSTEM

– Docker-Engine für Container

Mit Firmware für die Kommunikation mit der HPE Alletra Storage MP Plattform und einem Betriebssystem zur Handhabung minimaler Kernel-Änderungen und Pakete zusätzlich zu einem herkömmlichen Linux®-Betriebssystem befindet sich das „Gehirn“ des Systembetriebs in den Containern, die auf einem Container-

Orchestrator ausgeführt werden. Dank der Vorteile von Containern kann der Software-Stack jeder Funktion einen relativ unabhängigen Dienst zur Verfügung stellen. Das Resultat: mehr Effizienz, Konsistenz und Agilität.

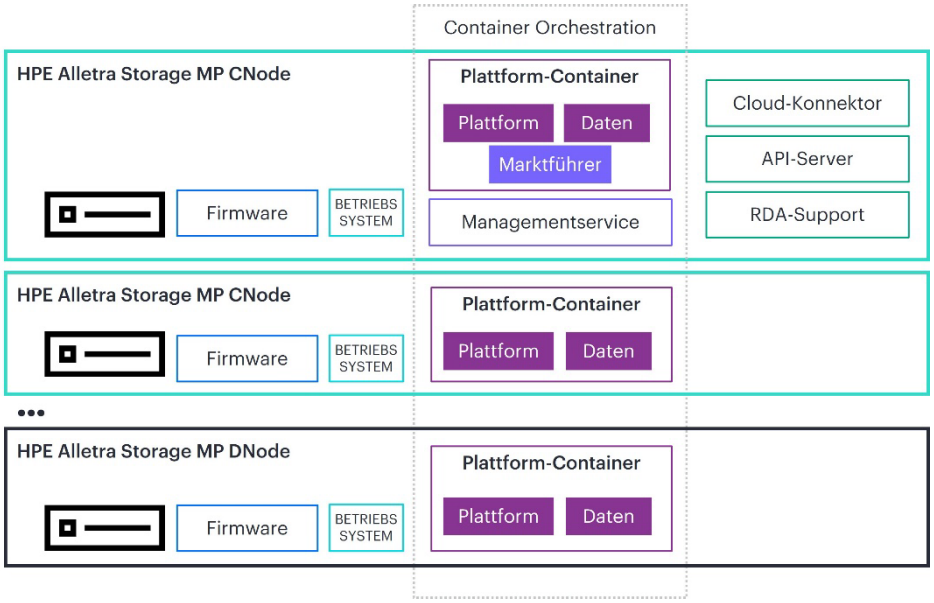


Abbildung 1. Blockdiagramm der Softwarearchitektur

Zwei Container (Abbildung 1) sind besonders zu beachten:

– Der Plattform-Container

– Der Managementdienst-Container

Der Plattform-Container wird sowohl auf jedem CNode als auch auf jedem DNode ausgeführt, allerdings verhalten sich diese Dienste auf CNodes und DNodes unterschiedlich. Der Plattform-Container auf den CNodes führt einen Großteil der Logik für die I/O-Vorgänge aus, dient auf den DNodes allerdings in erster Linie der Offenlegung des Datenspeichers für andere Dienste.

Ein Cluster verfügt stets nur über eine Instanz des Leader-Prozesses, und der Leader-Prozess wird auf einem CNode ausgeführt, um kritische Entscheidungen für den Cluster zu treffen, indem er beispielsweise bestimmt, auf welchem CNode der Managementdienst-Container ausgeführt wird. Durch den Einzelinstanz-Mechanismus des Leaders wird Komplexität bei Abstimmungen und Konsistenz vermieden. Andere Prozesse im Plattform-Container führen wichtige Funktionen zur Bedienung der I/Os für Dateidienste aus, beispielsweise Erasure Coding. Der Managementdienst-Container kontrolliert Managementvorgänge, beispielsweise die Aufnahme von Snapshots, die lokale Managementschnittstelle, Upgrades usw.

Darüber hinaus müssen einzige zusätzliche Prozesse beachtet werden:

– Der Cloud-Connector, der die gesamte Kommunikation über eine Tunnelverbindung zur HPE GreenLake Cloud herstellt und aufrechterhält

– Ein API-Server

– Ein RDA-Dienst für den Remote-Support

Dateisystem, Protokolle und protokollübergreifende Funktionen

Das Dateisystem ist so konzipiert, dass es die Attribute unterschiedlicher Dateisysteme verallgemeinert und zu einer einzelnen Datenbaumstruktur kombiniert, die als **Elementspeicher** bezeichnet wird. Alle Dateivorgänge über verschiedene Protokolle wie NFS und SMB werden Vorgängen im Elementspeicher zugeordnet. Dieses Konzept ermöglicht nicht nur eine mühelose protokollübergreifende Funktionalität (weil die zugrundeliegende

Datenstruktur gemeinsam genutzt wird und sich im gleichen Namespace befindet), sondern bietet auch für die Metadaten aller Protokolle dasselbe Maß an Datensicherheit, Datenreduktion und Transparenz.

Dank der zentralen Datenstruktur für den gesamten Speicherplatz (auch als „globaler Namespace" bezeichnet) werden keine Daten isoliert. Unterschiedliche Protokolle können verschiedene oder sich überlagernde Abschnitte des Elementspeichers offenlegen, indem sie die protokollspezifischen Einstellungen (Ansichtsrichtlinien) an einen Verzeichnispfad anhängen. Der angehängte Richtlinienabschnitt wird zu einer mountfähigen Dateifreigabe, die auch **Ansicht** genannt wird (siehe Abbildung 2).

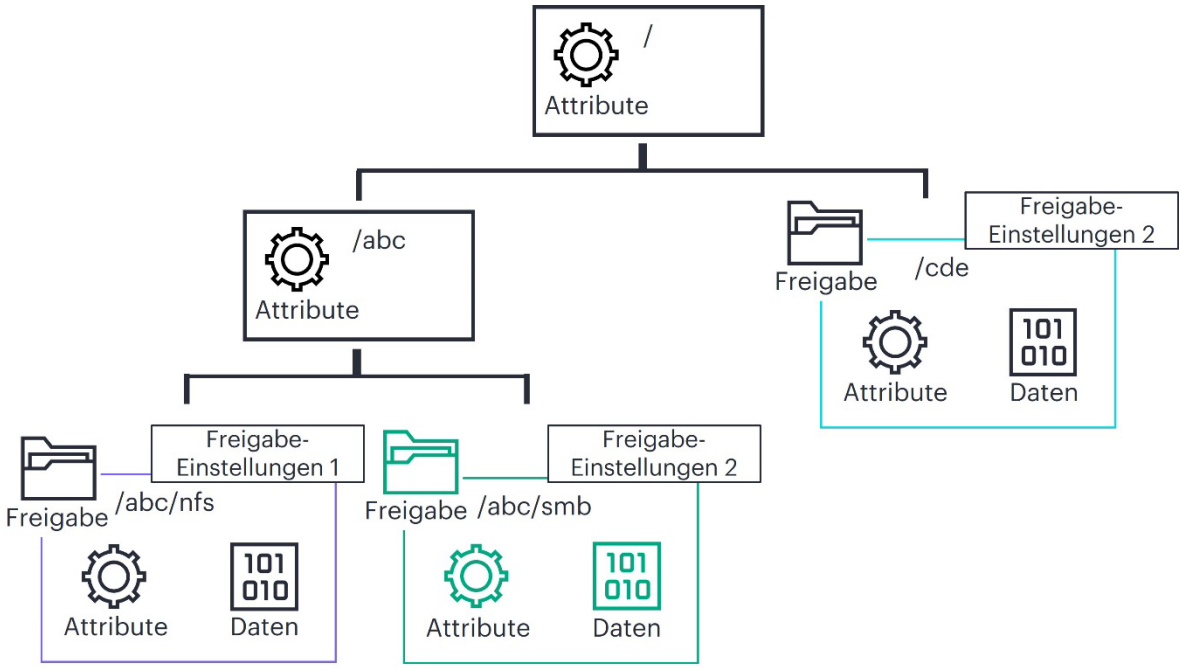


Abbildung 2. Beispiel für die Baumstruktur eines Elementspeichers (Dateisystem)

Einige wichtige Regeln für die Dateifreigabe können in den Freigabeeinstellungen definiert und anschließend mehreren Dateifreigaben zugeordnet werden. Diese wichtigen Regeln beinhalten unter anderem, welchen Protokollen die Festlegung von Berechtigungen für die Freigabe genehmigt wird (Sicherheitsfaktor), die Authentifizierungsquelle (Gruppenmitgliedschaftsquelle) sowie **VIP-Pools (virtuelle IP)**, die den Freigaben als auswählbare Access Points zur Verfügung stehen, sowie andere protokollspezifische Einstellungen.

Die Menge der Dateien oder Verzeichnisse wird durch die Anzahl der DBoxes und SSD-Laufwerkstypen begrenzt. In der folgenden Tabelle ist die Anzahl der unterstützten Dateien für jede von uns unterstützte DBox-Variante aufgeführt.

Tabelle 3. Tabelle mit der maximalen Dateianzahl pro Konfiguration

Konfiguration	SSD-basierte DBox mit 7 TB	SSD-basierte DBox mit 15 TB	SSD-basierte DBox mit 30 TB	SSD-basierte DBox mit 60 TB
Standarddichte (Dateien pro DBox in Millionen)	760	1.520	2.280	n/a
Standarddichte mit aktivierter Option „viele Dateien" (Dateien pro DBox in Millionen)	1.442	2.850	4.180	n/a
Hohe Dichte (Dateien pro DBox in Millionen)	n/a	1.900	n/a	5.130

Konfiguration	SSD-basierte DBox mit 7 TB	SSD-basierte DBox mit 15 TB	SSD-basierte DBox mit 30 TB	SSD-basierte DBox mit 60 TB
Hohe Dichte mit aktivierter Option „vielen Dateien“ (Dateien pro DBox in Millionen)	n/a	3.800	n/a	5.130

Die Grenze der Datei- und Verzeichnismenge steigt linear mit der Anzahl der DBoxes im System. Für Workloads, die eine Unterstützung von deutlich mehr Dateien erfordern, besteht die Möglichkeit, eine Funktion namens „Viele-Dateien-Unterstützung“ zu aktivieren, die die Dateianzahl der besagten Konfiguration verdoppelt. Es ist wichtig zu beachten, dass die Unterstützung vieler Dateien während der Installation aktiviert wird.

Globaler Namespace

Der globale Namespace, eine neu eingeführte Funktion in v3.1, ermöglicht einen pfadbasierten Zugriff auf Daten über mehrere Cluster hinweg, sowohl einen Lese- als auch einen Schreibzugriff. Dadurch wird eine hohe Einheitlichkeit beim Lesen und Schreiben gewährleistet. Ein Ursprungscluster im Kontext der globalen Namespace-Funktion bezieht sich auf den Cluster, in dem der Datenpfad erstellt wird, und Satellitencluster sind Cluster, in denen der im globalen Namespace konfigurierte Datenpfad auch zum Lesen und Schreiben verfügbar ist.

Dies ist ein typischer I/O-Workflow für den Datenzugriff auf einen im globalen Namespace konfigurierten Pfad. Zunächst wird auf dem Ursprungscluster eine Dateifreigabe erstellt. Auf die Daten können Clients dann über einen regulären Datenpfad zugreifen.

1. Wenn die globale Namespace-Funktion auf demselben Datenpfad konfiguriert ist, stehen dieselben Daten auch auf einem Satellitencluster für den Lese-/Schreibzugriff der Clients zur Verfügung.
2. Wenn ein Client Zugriff auf eine Datei in der für den globalen Namespace aktivierten Dateifreigabe auf einem Satellitencluster anfordert, erhält der Satellitencluster eine Kopie der Datei und eine Lese-Lease vom Ursprungscluster. Beim ersten Zugriff auf die Datei dauert es normalerweise länger (abhängig von der Netzwerklatenz zwischen dem Ursprungscluster und dem Satellitencluster). Sobald jedoch die erste Kopie der Datei auf dem Satellitencluster abgerufen wurde, sind alle nachfolgenden Lesevorgänge des Inhalts der Dateifreigabe viel schneller, da auf den Inhalt direkt aus der zwischengespeicherten Version der Datei auf dem lokalen Satellitencluster zugegriffen wird, solange die Lese-Lease gültig ist.
3. Entweder nachdem eine konfigurierbare feste Zeitspanne erreicht ist oder wenn ein Schreib-I/O für die Datei(en) ausgegeben wird, läuft die Lese-Lease ab und der zwischengespeicherte Inhalt dieser Datei(en) wird ungültig.

Wenn Sie die Verwendung globaler Namespaces in Erwägung ziehen, ist es wichtig, sich über deren Einschränkungen im Klaren zu sein. Replikation und globaler Namespace können nicht zusammen auf demselben Datenpfad verwendet werden. Der globale Namespace ist nur für den NFSv3-Zugriff auf Satellitenclustern verfügbar. Die Dateifreigaben mit aktiviertem globalen Namespace können jedoch weiterhin für andere protokollübergreifende Zugriffe auf dem Ursprungscluster verfügbar sein. Die Schreibleistung kann durch die Latenz zwischen dem Ursprungscluster und den Satellitenclustern beeinträchtigt werden. Diese Funktion gilt noch nicht als Funktion zur Verbesserung des Datenschutzes, da die Daten nur auf dem Ursprungscluster verfügbar gemacht werden. Insbesondere wenn der Ursprungscluster nicht erreichbar ist, stellen die Satellitencluster keine Daten für diesen Pfad bereit. Wenn ein Satellitencluster nicht erreichbar ist, funktionieren der Ursprungscluster und die anderen Satellitencluster normal.

Leistung

Die HPE GreenLake for File Storage Lösung liefert hohe Leistung, die in beispiellosem Maße vorhersehbar, zuverlässig und skalierbar ist. Ebenso bietet sie lineare Leistung im Zuge der Skalierung – unabhängig vom

Computing für die Leistung oder vom Datenspeicher für die Kapazität und ohne pauschale Overheads. Da zwei Probleme traditioneller, auf einer Shared-Nothing-Architektur basierender Systeme bewältigt werden, ist dieses Konzept aus zwei wichtigen Gründen einzigartig:

– Die architektonischen Einschränkungen traditioneller Systeme sind mit einem pauschalen Overhead verbunden, da das System ab einem bestimmten Punkt zu komplex für ein effektives Management und somit instabil wird.

– Diese architektonischen Einschränkungen erzwingen die Skalierung in Kombination mit Computing und Datenspeicher.

In den nachfolgenden Abschnitten werden die wichtigsten Elemente der Architektur von HPE GreenLake for File Storage mit einem Schwerpunkt auf Leistung und Skalierbarkeit beschrieben.

Datenpfad

Der Datenpfad eines Schreib-I/O-Vorgangs ist im Vergleich zu anderen NAS-Lösungen (Network Attached Storage) einzigartig, da er so konzipiert ist, dass die üblichen Herausforderungen von NAS im Zusammenhang mit der Erzielung von Datenkonsistenz mit flüchtigen Speichern überwunden werden. Abbildung 3 illustriert den Schreib-Datenpfad.

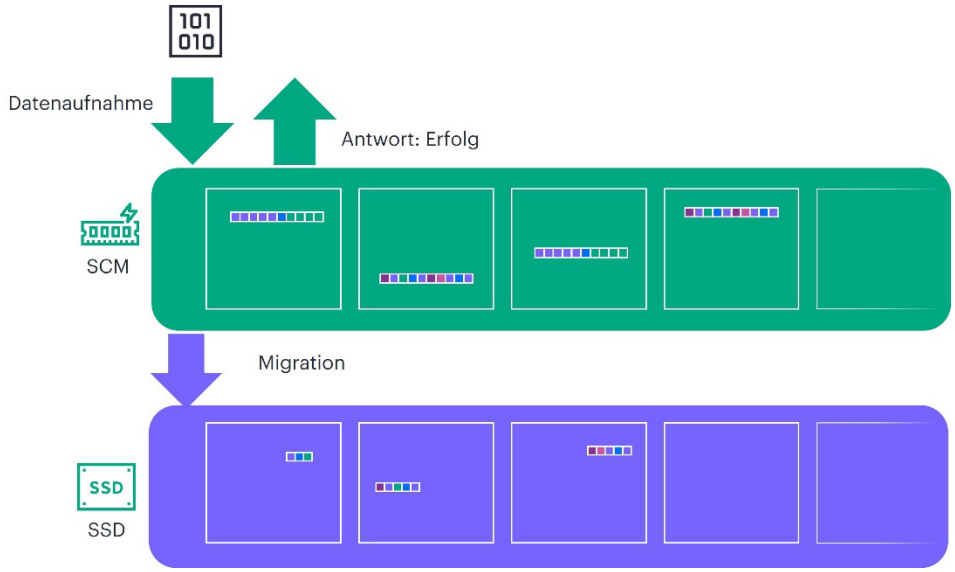


Abbildung 3. Schreib-Datenpfad

Da der Schreib-I/O von einem Client über ein beliebiges herkömmliches NAS-Protokoll stammt und beliebige CNodes über einen der VIP-Pools von Mountpunkt-IPs erreicht, kann der CNode die Schreibanfrage nicht als erfolgreich abgeschlossen bestätigen, bis die folgenden drei Aufgaben durchgeführt wurden:

1. Die Daten werden auf mehrere SCM-Laufwerke geschrieben (Schreibpuffer-RAID oder gespiegelt), sodass bei einem Ausfall des SCM-Laufwerks keine Daten verloren gehen.
2. Daten werden zufällig auf mehrere SCM-Laufwerke aufgeteilt, um den Durchsatz zu erhöhen und Konflikte zu minimieren.
3. Metadaten werden im Sinne der Konsistenz in der internen Datenstruktur (Baum) aktualisiert.

Mit diesem Prozess wird die Konsistenz der Daten gewährleistet. Nach dem Abschluss des Schreib-I/O-Vorgangs befinden sich die geschriebenen Daten auf persistentem Datenspeicher (SCM) mit Redundanz. So ist die Wahrscheinlichkeit eines Datenverlustes äußerst gering.

Zusätzlich zu den unmittelbaren Aufgaben zum Abschließen eines Schreib-I/O-Vorgangs müssen die Daten auch vom SCM auf ein SSD-Laufwerk migriert werden. Bei dieser Migration handelt es sich um einen Out-of-Band-Prozess. Er wird von allen CNodes im Cluster in seinem eigenen Tempo durchgeführt. Die Migration erfolgt meist dann, wenn die Clients inaktiv oder schwach ausgelastet sind. Im Zuge der Befüllung des SCM wird dieser Prozess immer aggressiver. Allerdings können die CNodes bei Bedarf die Antworten auf die Schreibanfragen des Clients drosseln, um den Schreibdurchsatz zu verlangsamen, damit die Migration mit dem Tempo der Datenaufnahme mithalten kann. Dieses Problem kann durch eine Erweiterung der DBoxes, bei der der verteilte SCM-Layer vergrößert wird, gelöst werden. Daten bleiben immer im SCM erhalten, bis sie sicher zur SSD migriert werden können. Die Gesamtkosten eines Schreib-I/O-Vorgangs (einschließlich der Kosten der Migration zuzüglich zu den ursprünglichen Kosten eines Schreib-I/O-Vorgangs auf SCM) umfassen die folgenden Schritte:

1. Daten werden im SCM geschrieben (Schreibpuffer-RAID oder gespiegelt).

   Beachten Sie, dass sämtliche nachfolgenden Schritte erst während der Migration von SCM auf SSD durchgeführt werden, nachdem die Schreibanfragen den Clients bereits als erfolgreich bestätigt wurden.
2. Daten werden in den RAM des DNodes eingelesen.
3. Die CNode-CPU wechselt zum Migrationsprozess, der die Berechnung der Datenreduktion usw. beinhaltet.
4. Die Daten werden einmal auf die SSD geschrieben.

Der Lese-Datenpfad ist einfacher als der Schreib-Datenpfad. Nachdem ein CNode über ein beliebiges NAS-Protokoll eine Leseanfrage von einem Client erhält, ermittelt er den Speicherort der Daten über die SCM-Metadaten. Dann holt der CNode die Daten vom SCM bzw. der SSD ab und gibt sie an den Client aus. Da CNodes den gleichen Zugriff auf SCM und SSD haben, hat es keine Auswirkungen auf die Leistung des Lese-I/O-Vorgangs, ob die angeforderten Daten sich auf dem SCM oder der SSD befinden.

Controller-Pooling, Servicequalität und Mandantenfähigkeit

HPE GreenLake for File Storage bietet eine leistungsstarke Funktion namens **Controller-Pooling**. Damit können Sie entweder alle oder nur bestimmte Computing-Nodes einer Anwendung zuweisen, um eine ausreichende Rechenleistung für diese Anwendung zu gewährleisten. Hierzu werden VIP-Pools erstellt, die entweder auf alle oder nur auf bestimmte CNodes verteilt werden. Stellen Sie sich zum Beispiel eine Anwendung vor, die alle Ressourcen erfordert, die sie bekommen kann, um ihre SLA zu erfüllen. Dieser Anwendung wird ein VIP-Pool zugewiesen, dem 80–90 % der CNodes zugewiesen sind. Die verbleibenden 10–20 % der CNodes können Mandanten mit niedriger Priorität zugewiesen werden, beispielsweise Sicherungen oder Replikation. Sollte die Anwendung allerdings je alle verfügbaren Computing-Ressourcen benötigen, könnte ihr VIP-Pool flexibel und ohne Unterbrechungen oder langwierige Abstimmung auf alle Nodes erweitert werden. Fällt ein CNode aus, werden alle virtuellen IPs zum verbleibenden CNode verschoben.

Bei dieser Lösung stellen die CNodes den Front-End-Protokolldienst für die Clients oder Anwendungen bereit, und die gesamte Intelligenz, beispielsweise das Striping der Daten auf die DBoxes oder das Lesen der Daten von den DBoxes, wird auf diesen CNodes ausgeführt. Dennoch bleiben CNodes flüchtig. Das bedeutet, dass sie innerhalb dieser Architektur keinen Status enthalten, wie im Abschnitt DASE-Architektur erläutert.

QoS-Richtlinien (Quality of Service – Servicequalität) ermöglichen Ihnen die Definition einer fein abgestimmten Kontrolle von Bandbreite und IOPS nach einzelnen Freigaben oder pro Benutzer mit Mindest- und Höchstgrenzen. Zusätzlich zu den minimalen und maximalen QoS-Grenzwerten kann die QoS-Richtlinie ein Burst-Limit und ein Credit-Limit durchsetzen. Wenn ein Burst-Limit festgelegt ist, werden Burst-Credits gesammelt, wenn die Workload weniger Ressourcen verbraucht als durch das maximale Limit festgelegt. Die Credits können später ausgegeben werden, um eine Leistung zu erzielen, die den Höchstwert überschreitet, bis hin zum konfigurierten Burst-Limit. Der Betrag der Credits ist kumulierbar und durch ein Credit-Limit begrenzt. Wenn beispielsweise eine QoS-Richtlinie ein maximales Limit von 100 MB/s, ein Burst-Limit von 1.000 MB/s und ein Credit-Limit von 10.000 definiert, wird das Credit-Limit nach 100 Sekunden Leerlaufzeit erreicht. Mit diesem Credit-Guthaben kann die

Anwendung 10 Sekunden lang mit 1.000 MB/s laufen. Anschließend wird die Geschwindigkeit auf 100 MB/s gedrosselt.

Die QoS-pro-Ansicht-Richtlinie kann einen QoS-Bereitstellungsmodus auswählen und statische und/oder kapazitätsbasierte QoS-Grenzwerte festlegen. Mit QoS können jedoch in der Benutzeransicht statische QoS-Grenzwerte für einen oder mehrere Benutzer festgelegt werden.

Bei dieser Architektur mit globalem Namespace werden Daten mit einem oder mehreren Protokollen, die durch die Freigaberichtlinien definiert und verwaltet werden, für die Clients offengelegt. Diese Richtlinien und Freigaben können mit einem Konstrukt verknüpft werden, das als **Mandant** bezeichnet wird und die Festlegung von Regeln zur Trennung von Freigaben und Freigaberichtlinien und damit letztendlich die **Mandantenfähigkeit** ermöglicht. Ein Mandant kann über seinen eigenen IP-Adressbereich für den Zugriff auf die mit ihm verknüpften Freigaben verfügen. So kann jeder Mandant seine eigenen Identitätsanbieter haben, beispielsweise Active Directory, LDAP oder NIS (nicht mehr als acht Active Directory-Anbieter, acht LDAP-Anbieter und ein NIS-Anbieter pro System). Jeder Mandant kann über seine eigenen Verschlüsselungsschlüssel für die Verschlüsselung ruhender Daten verfügen.

Controller-Pooling, QoS und Mandantenfähigkeit sind voneinander unabhängig, können allerdings entweder separat oder gemeinsam genutzt werden. Im Wesentlichen ermöglicht **Controller-Pooling** eine grobe Kontrolle über die Nutzung der CNode-Ressourcen und der Bandbreite. **Mandantenfähigkeit** ermöglicht eine fein abgestimmte Kontrolle über die Ansicht und den Abruf von Daten und **QoS** ermöglicht eine fein abgestimmte Kontrolle über die Ressourcennutzung.

Höhere NFS-Leistung

Die HPE GreenLake for File Storage Lösung ermöglicht eine höhere NFS-Leistung (Abbildung 4). Zusätzlich zum herkömmlichen NFS-Mount mit Single-Socket-Verbindung zwischen Client und Datenspeicher-Port steht eine Multipathing-Funktion für den Multi-Socket-Zugriff über Linux nConnect zur Verfügung. Mit nConnect kann bei bestimmten Linux-Distributionen wie RHEL 8.3 oder Ubuntu 20.04 oder neuer die nConnect-Mountoption zur Konfiguration von bis zu 16 TCP-Verbindungen zwischen dem Client und dem einzelnen Datenspeicher-Port (VIP-Pools auf dem Datei-Cluster) verwendet werden.

Darüber hinaus unterstützt die Lösung Multipathing für NFS-Clients. Mit dieser Funktion kann ein Client mehrere Verbindungen von mehreren Ports zu mehreren Adressen öffnen. Dies wird für die Versionen NFSv3 und NVSv4.1 unterstützt und erfordert die Installation eines Pakets auf dem NFS-Client.

Zusätzlich unterstützt die Lösung NFS over RDMA. Dabei werden CPU-Ressourcen umgangen, um eine bessere Leistung zu erzielen. Und mit GPUDirect Storage™ erreicht die Leistung ein völlig neues Niveau, da sogar der Arbeitsspeicher auf der Client-Seite umgangen wird, um eine bemerkenswerte Leistung von 170 GB/Sek. pro Host zu erzielen. So erzielt die Lösung dank mehrerer Verbindungen und Pfade eine bessere Skalierung der NFS-Leistung.

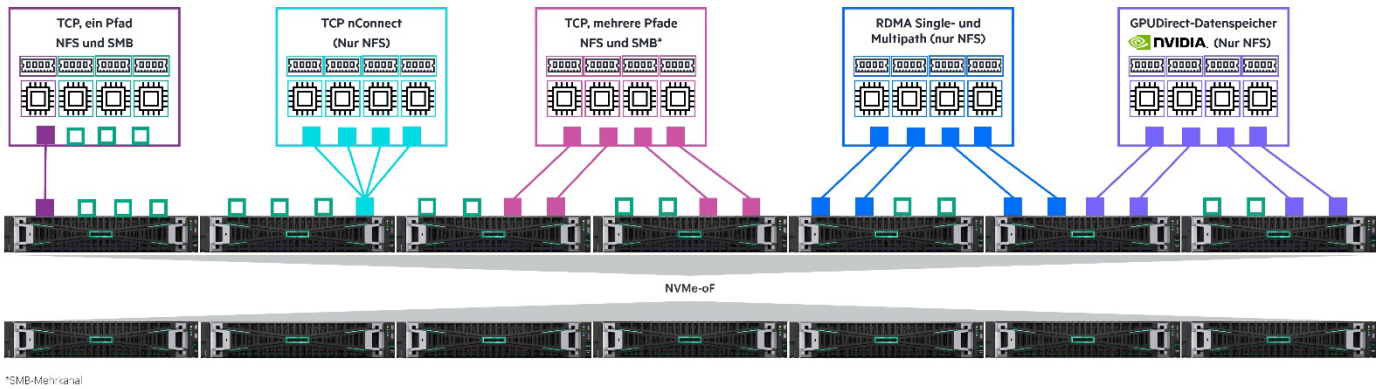


Abbildung 4. Anwendungsfälle für verbesserte Leistung

Datenspeicherkapazität

Die HPE GreenLake for File Storage Lösung bietet eine Nutzkapazität ab ca. 220 TB, die bei Bedarf erheblich ausgebaut werden kann. Die Ausgangskonfiguration enthält ein 100-GbE-Switch-Paar (Leaf) und lässt sich zu einem größeren System mit Spine und Leafs für die NVMe-Fabric am Back-End skalieren. SCM-Laufwerke werden bei der Gesamtkapazität nicht berücksichtigt. Es ist wichtig, die Begriffe und Definitionen im Zusammenhang mit der Kapazität zu verstehen, insbesondere unter Berücksichtigung der Datenreduktion:

- **Rohkapazität** bezeichnet die aggregierte Gesamtkapazität aller NVMe-Laufwerke ohne Berücksichtigung von Speicherplatzreservierung, Erasure Coding oder Datenreduktion.
- **Logische Kapazität** steht für die auf dem Array gespeicherten Datenanwendungen.
- **Nutzkapazität** ist die verfügbare Kapazität nach der Berechnung von Erasure Coding und Speicherplatzreservierung. Datenreduktionsberechnungen werden darin nicht berücksichtigt.
- **Effektive Kapazität** ist die Kapazität, bei der sämtliche Datenreduktionsberechnungen auf dem Array berücksichtigt sind.
- **Zusatzkapazität** ist die Kapazität, die nach der Löschung von Daten vom Client wieder für die Nutzkapazität freigegeben wird. Zugrundeliegende Snapshots der gelöschten Daten verhindern die Löschung des zusätzlichen Speicherplatzes.

Datenreduzierungstechnologien

Bei der Datenreduktion mit HPE GreenLake for File Storage wurden in vielerlei Hinsicht bedeutende Fortschritte erzielt. In den vergangenen Jahrzehnten boten die meisten Datenspeicherlösungen eine Deduplizierung und verlustfreie Komprimierung von Datensätzen innerhalb des Datenspeicher-Subsystems. HPE GreenLake for File Storage bietet einen einzigartigen Datenreduktionsansatz und wendet diese Techniken im gesamten globalen Namespace an, um erstklassige Datenreduktionsverhältnisse zu erzielen.

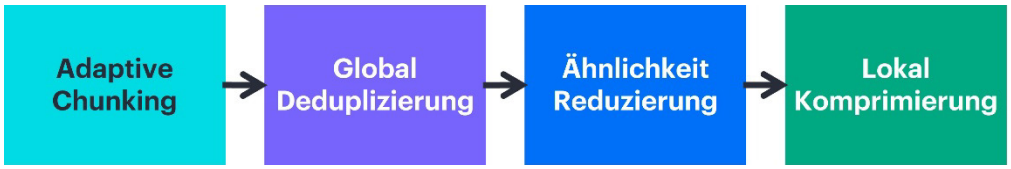


Abbildung 5. Datenreduktions-Workflow

Die Datenreduktion ist bei HPE GreenLake for File Storage ein Inline-Prozess (Abbildung 5). Während der Datenmigration (häufig als „De-Staging" bezeichnet) von SCM auf SSD durchläuft die Daten-Pipeline einen Prozess namens **adaptives Chunking**. Anschließend vergleicht die globale Deduplizierungs-Engine die Datenblöcke im gesamten Namespace. Auf dieser Stufe ermittelt der Prozess der ähnlichkeitsbasierten Reduktion Datenblöcke, die sich ähneln, aber nicht identisch sind, um die Größe der Daten noch weiter zu reduzieren. Am Ende findet die lokale Komprimierung statt. Auch wenn es sich bei den verwendeten Datenreduktionstechnologien um blockbasierte Reduktionsmethoden handelt, gibt es im Gegensatz zum traditionellen Block Storage keine feste Blockgröße.

Adaptives Chunking der Blockgröße

Das optionale, standardmäßig aktivierte adaptive Chunking der Blockgröße trägt mithilfe einer gleitenden Blockgröße dazu bei, die Wahrscheinlichkeit der Durchführung einer ähnlichkeitsbasierten Reduktion und Deduplizierung zu maximieren. Bei einer adaptiven Blockgröße zwischen 16 und 64 kB passt der Algorithmus die Blockgröße anhand der Wahrscheinlichkeit einer bestmöglichen Datenreduktion an. Somit hat eine Datei, die bei einem späteren Status bearbeitet oder geändert wird, nur minimale Auswirkungen auf den Blocksatz der Datei. Dennoch wird im Gegensatz zum Chunking mit fester Blockgröße, bei dem alle Dateiblöcke nach einem Bearbeitungsvorgang nicht mehr mit ihrem ursprünglichen Status übereinstimmen, die größtmögliche Wahrscheinlichkeit einer Datenreduktion erzielt.

Globale Deduplizierung

Jeder Datenblock generiert einen kryptografischen Hashwert, und die Hashwerte aller Daten auf dem Laufwerk werden auf SCM gespeichert. Bevor die Daten im Migrationsprozess auf den NVMe-SSD-Layer geschrieben werden und nachdem die Daten das adaptive Block-Chunking durchlaufen haben, werden die Daten mit allen vorhandenen Hashwerten im gesamten Namespace abgeglichen. Wird eine Übereinstimmung gefunden, werden die Daten als Verweis auf den duplizierten Block gespeichert, der in den vorhandenen Daten gefunden wurde. Dieser Deduplizierungsprozess kann erheblich effektiver ausfallen als die lokale Deduplizierung, da die Shared-Everything-Architektur allen CNodes den Zugriff auf den gesamten Datensatz ermöglicht – und zwar ohne dass der Ost-West-Datenverkehr CPU-Zyklen in Anspruch nimmt.

Ähnlichkeitsbasierte Reduktion

Statt eine genaue Übereinstimmung für die Deduplizierung eines Blocks zu finden, vergleicht die ähnlichkeitsbasierte Reduktion einen berechneten Ähnlichkeits-Hashwert, um Blöcke mit hoher Ähnlichkeit, aber geringfügigen Unterschieden zu finden. Wenn sich zwei Blöcke anhand eines kollidierenden Ähnlichkeits-Hashwerts inhaltlich in ausreichendem Maße ähnlich sind, werden sie als Verweis auf einen dedupizierten Baseline-Block und Deltas auf Byte-Ebene gespeichert (Abbildung 6).

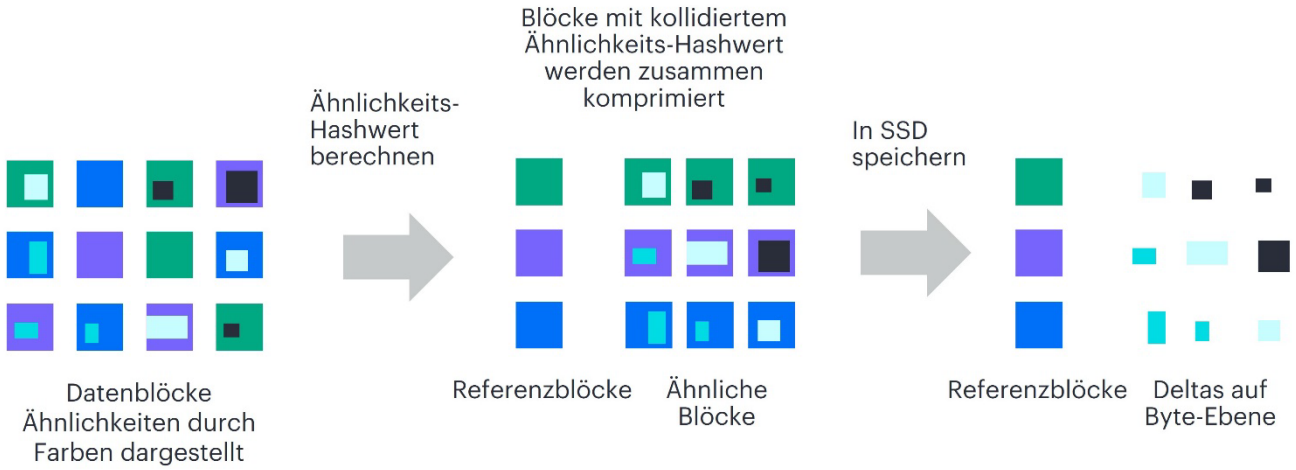


Abbildung 6. Ähnlichkeitsbasierte Reduktion

Lokale Komprimierung

Zstandard (ZSTD), ein verlustfreier Komprimierungsalgorithmus, wird zur Komprimierung von Datenblöcken verwendet, um die Blockgröße zu reduzieren. ZSTD ist ein schneller und verlustfreier Komprimierungsalgorithmus speziell für Echtzeit-Komprimierungsszenarien und bessere Komprimierungsverhältnisse.

Defragmentierung

Dank der höheren Stripe-Größe für höhere Leistung kann ein einzelner Stripe mehrere Dateien enthalten. Daher gelten Teile der Stripes als veraltet, wenn Daten aktualisiert oder gelöscht werden. In einem Defragmentierungsprozess werden die veralteten Teile dieser Stripes zurückgewonnen, wenn ein Stripe genügend veralteten Platz aufweist, um die Kosten des Defragmentierungsprozesses zu rechtfertigen. Der Defragmentierungsprozess findet nur bei Bedarf statt. In anderen Worten: Je stärker die Kapazität ausgelastet ist, desto aggressiver fällt die Defragmentierung aus. Dieser Prozess wirkt sich stärker auf die Leistung aus, da die Kapazität nahezu vollständig belegt wird und der Prozess mehr CPU-Zyklen von anderen Vorgängen abzieht.

Kontingente

Bei einer Lösung wie HPE GreenLake for File Storage, die einen globalen Namespace bietet, wäre ein Mechanismus zur granularen Zuweisung von Kapazitätsgrenzen trivial. An dieser Stelle kommen Kontingente zum Tragen. Kontingente können auf Verzeichnisebene nach Kapazität und/oder der Menge der Dateien und Verzeichnisse, die unter einer Freigabe oder einem Verzeichnis erstellt werden können, festgelegt werden. Darüber hinaus können Kontingente auch nach Benutzer und nach Gruppe festgelegt werden, um die Kapazität,

die einzelne Benutzer und Gruppen innerhalb eines Verzeichnisses nutzen können, sowie auch die Menge von Dateien und Verzeichnissen einzuschränken.

Datenresilienz und Zuverlässigkeit

Datenresilienz ist bei HPE GreenLake for File Storage auf jeder Ebene integriert, insbesondere beim Schreib-Datenpfad, der Hardware-Fehlertoleranz und der Disaster Recovery. In diesem Abschnitt des Whitepapers werden die technischen Einzelheiten hinter dem Schutz von Daten durch diese Funktionen behandelt.

Schreib-Datenpfad

Zwei wesentliche Funktionen, die in den Schreibpfad zum SCM-Layer integriert sind, helfen bei der Maximierung der Datenresilienz: **Schreibpuffer-RAID**[1] oder **Spiegelung** und **zufälliges Sharding**, je nach Systemkonfiguration. Alle SCM-Geräte sind in 8 GB große Abschnitte unterteilt. Wenn ein eingehender I/O-Vorgang einen CNode erreicht, fordert er eine Abschnittsadresse an und erhält die physischen Adressen von zwei SCM-Geräten, auf die Daten geschrieben werden sollen. Alle Abschnitte werden je nach Systemkonfiguration entweder im RAID 6-Format geschrieben oder gespiegelt. Diese Abschnittspaare werden zufällig und gleichmäßig auf die verfügbaren physischen SCM-Geräte verteilt. Die CNodes schreiben direkt auf diese SCM-Geräte und bestätigen dem Client die Schreibanfrage als erfolgreich, nachdem sie die Bestätigung erhalten haben, dass Daten dauerhaft an beiden Orten gespeichert werden. Die Daten-Chunks in diesen Abschnitten können unterschiedliche Größen aufweisen, und Daten können auf jeden logischen Adressabstand der 8-GB-Abschnitte geschrieben werden.

Festgeschriebene Daten befinden sich auf keiner Form von nichtflüchtigem Speicher, sodass das System bei einem Ausfall von der Bereinigung von jeglichem „unsauberem" Cache befreit wird. Nachdem die Daten auf den SCM geschrieben wurden, ist die Wahrscheinlichkeit eines Datenverlustes äußerst gering.

Schutz durch lokal decodierbares Erasure Coding

Erasure Coding von SSDs kann durch Wide Striping und Paritätsredundanz vereinfacht ausgedrückt werden. Die Breite eines Erasure–Coding-Strips ist die Menge der Laufwerke, auf die ein Stripe verteilt ist, und der Wert wird als **D + P** definiert. Die Parität (P) ist immer 4. Das bedeutet, dass maximal 4 SSD-Laufwerke ohne Datenverlust gleichzeitig ausfallen können, und für die Daten (D) gilt die folgende Approximation: Nehmen Sie die Anzahl der intakten SSD-Laufwerke und ziehen Sie 4 ab. D ist der kleinere Wert aus 146 und der im vorherigen Schritt berechneten Zahl. Der Overhead-Anteil kann durch die Berechnung von P/D approximiert werden. Das Wide Striping verringert den Overhead-Anteil und ist daher bei einer großen Konfiguration von Vorteil. Größere Konfigurationen können dabei von einem Overhead von gerade einmal 2,7 % profitieren.

Das Konzept ohne feste Gruppierungsgröße ermöglicht eine einfache Integration verschiedener Laufwerksgrößen und Erweiterung sowie eine unkomplizierte Handhabung von Laufwerksausfällen. Im Rahmen der Erweiterung der Lösung (beispielsweise wenn weitere DBoxes zur bestehenden Konfiguration hinzugefügt werden) bestimmt die neue Konfiguration die neue, größere Stripe-Größe, und die vorher geschriebenen Stripes mit der geringeren Stripe-Größe werden allmählich kopiert und zur neuen Stripe-Größe für die gesamte neue Konfiguration umgeschrieben.

Obwohl der in dieser Lösung verwendete Erasure-Coding-Algorithmus mit Wide Striping eine starre Datenresilienz ermöglicht, bietet der lokal decodierbare Aspekt des Algorithmus zusätzlich den Vorteil, dass ein beschädigter Teil des Stripes rekonstruiert werden kann, ohne die gesamte Stripe-Breite lesen zu müssen. Tatsächlich kann **1/X** der Daten-Stripes lokal decodierbar sein, wobei X für die Anzahl der Schutz-Stripes im Satz steht. Bei einer typischen **nD+4P**-Codierung muss nur **1/4** der Daten im Zuge einer Rekonstruktion gelesen werden.

Hardware-Fehlertoleranz

In den meisten Fällen werden mehrfache Ausfälle auf allen Ebenen der Hardwarekomponenten mit nur wenigen Ausnahmen toleriert. Die folgende Liste enthält die Details zu den einzelnen Hardwarekomponenten:

[1] Die Schreibpuffer-RAID-Funktion gilt derzeit nur für die neu installierte Variante von HPE GreenLake for File Storage mit hoher Dichte.

- **SCM:** Im Allgemeinen kann der Ausfall eines SCM-Geräts ohne Datenverlust oder Unterbrechung der Bedienung von I/Os toleriert werden. Beim Ausfall eines SCM-Geräts weist der Leader-Prozess alle CNodes an, alle betroffenen 8-GB-Abschnitte des SCM-Geräts parallel zu rekonstruieren. Währenddessen werden alle Leseanforderungen durch Lesen vom Replikat erfüllt. Obwohl diese 8-GB-Abschnitte gleichmäßig verteilt sind und der Rekonstruktionsdatenverkehr hochgradig parallel verläuft, kommt es während der SCM-Rekonstruktion dennoch zu Leistungseinbußen. Im typischen Produktionseinsatz dauert die SCM-Rekonstruktion höchstens 30 Minuten, und im Worst-Case-Szenario, in dem die SCM-Geräte vollständig ausgelastet sind, kann die SCM-Rekonstruktion bis zu einer Stunde dauern. Wenn das zweite SCM-Gerät nicht ausfällt, bevor die SCM-Rekonstruktion abgeschlossen ist, entstehen keine Datenverluste. Allerdings kann der Ausfall von mehr als einem SCM-Gerät zu einem Datenverlust führen.
- **SSD:** Beim Ausfall eines einzelnen SSD-Laufwerks oder bei bis zu drei gleichzeitig ausfallenden SSD-Laufwerken ermöglicht das lokal decodierbare Erasure Coding durch das Lesen von nur 1/4 des Datensatzes des Stripes die Wiederherstellung des Stripe-Abschnitts von den beschädigten SSD-Laufwerken. Während der Rekonstruktion kommt es zu keinen Leistungseinbußen am Front-End. Fallen jedoch vier SSD-Laufwerke gleichzeitig aus, muss das System 100 % der Blöcke vom Stripe lesen, um die beschädigten Daten von den ausgefallenen Laufwerken zu rekonstruieren. Während dieser Zeit kommt es zu Leistungseinbußen, da die Datenrekonstruktion priorisiert wird, um dazu beizutragen, die Wahrscheinlichkeit eines Datenverlustes zu minimieren.
- **CNode:** Die DASE-Architektur erlaubt eine unbegrenzte Anzahl von CNode-Ausfällen, solange ein CNode intakt bleibt. Alle CNodes haben den gleichen Zugriff auf den gesamten Speicherplatz und Metadatenpool und sind statusfrei/flüchtig. Intakte CNodes können die Workloads ohne I/O-Unterbrechungen fortsetzen, indem sie sich den VIP-Pooling-Mechanismus zunutze machen.
- **DNode:** DNodes werden in HA-Paaren betrieben und legen SSD und SCM für die CNodes offen. Jede SSD und jeder SCM unterstützt Duo-Ports, und Duo-Port-Laufwerke können während eines DNode-Ausfalls vom intakten Node im gleichen DBox-Gehäuse aufgerufen werden. Wenn während eines DNode-Ausfalls in jedem DBox-Gehäuse mindestens ein DNode funktionsfähig bleibt, kommt es zu keiner Unterbrechung für den Client. Die Leistung kann durch den Verlust von NICs und PCI-Bus-Durchsatz auf der ausgefallenen DBox beeinträchtigt werden.
- **CBox:** Die CBox hat die gleiche Fehlertoleranz wie der CNode. Das System bleibt funktionsfähig, solange mindestens ein CNode aus anderen CBoxes im Cluster intakt bleibt.
- **DBox:** Ein DBox-Ausfall ist äußerst unwahrscheinlich, da die DBox mit redundanten Komponenten konstruiert ist. In dem unwahrscheinlichen Fall, dass es zu einer Ausfallzeit kommt, tritt aufgrund eines DBox-Ausfalls kein Datenverlust auf. Um den Betrieb nach einem DBox-Ausfall wiederaufzunehmen, müssen alle Laufwerke aus dem ausgefallenen Gehäuse entfernt und in eine intakte Ersatz-DBox eingesetzt werden. In einem zukünftigen Release wird das System mit einer höheren Anzahl von DBoxes für DBox-Redundanzen konfiguriert werden können.
- **Datennetzwerk:** Jeder CNode und jede DBox bieten Redundanz mit Ports und Konnektivität zu den internen Switches sowie zu den TOR-Switches im Kundennetzwerk. Fallen genügend Ports oder NICs aus, um einen Node oder ein Gehäuse unerreichbar zu machen, kann dieser Ausfall wie ein Node- oder Enclosure-Ausfall behandelt werden.
- **Verwaltungsnetzwerk:** Ein Ausfall des Managementnetzwerks kann neben einer Unterbrechung des Managements (Managementdienst, SSH usw.) auch folgende Konsequenzen haben: kein Versand von Call-Home-Paketen, Verlust des Registry-Zugriffs, der wiederum eine Änderung der Benutzerpopulation und eine Unterbrechung des Cluster-Zugriffs zur Folge hat, sowie Verlust des Zugriffs auf das DNS, was zu einer Mount-Unterbrechung führen kann.

Online-Upgrades

Software-Upgrades werden in regelmäßigen Abständen online implementiert. Bei den meisten Konfigurationen ist einer von N Nodes während seines Upgrades zu einem bestimmten Zeitpunkt offline. Bei größeren Konfigurationen werden mehrere Nodes gleichzeitig aktualisiert. Grundsätzlich wird bei einem Upgrade der

Managementdienst-Container zuerst aktualisiert. Im Anschluss folgen die Software-Upgrades einer DBox nach der anderen. Anschließend wird die Software eines CNodes nach dem anderen aktualisiert. Davon ausgenommen ist der Hosting-Node des Managementdienstes, der als letzter aktualisiert wird.

Nicht jedes Upgrade erfordert einen Node-Neustart. Upgrades erfordern keine systemweite Downtime, und es kommt während eines Upgrades nur zu einer minimalen Leistungsbeeinträchtigung. Firmware-Upgrades erfordern keine Downtime des Systems. Die Dauer eines Upgrade-Prozesses variiert je nach Größe der Konfiguration. Bei einer Konfiguration von 12 bzw. der kleinsten Konfiguration aus einem einzelnen CNode und einer einzelnen DBox nimmt ein Software-Upgrade ohne Aktualisierung des Betriebssystems in der Regel etwa 30 Minuten in Anspruch.

Snapshots, Aufbewahrung und Unveränderlichkeit

Snapshots bieten eine effektive Methode für den lokalen Ransomware-Schutz. Bei der Aufnahme eines Snapshots von einem beliebigen Verzeichnisbaum wird eine konsistente Point-in-Time-Aufzeichnung der Daten unter diesem Pfad erstellt. Zum Zeitpunkt der Aufnahme des Snapshots kommt es in der Regel zu keinen Auswirkungen auf die Kapazitätsauslastung oder Leistung. Die Kapazitätskosten von Snapshots entstehen bei der Löschung der Daten, und die Auslastung wird durch die Zusatzkapazität abgebildet.

Bei der Löschung von Snapshots wird die Leistung kaum oder überhaupt nicht beeinträchtigt, wobei es zwei Ausnahmen gibt:

– Die Löschung eines Snapshots, der eine große Menge an gelöschten Daten enthält

– Die Löschung mehrerer Snapshots, die geringfügige Änderungen an der gleichen Datei enthalten

In beiden Fällen kann der Defragmentierungsprozess zur Bereinigung einer großen Menge an veralteter Kapazität bzw. eine zunehmende Menge an kleinen zufälligen I/O-Vorgängen während der Defragmentierung die Leistung vorübergehend beeinträchtigen.

Snapshots können entweder manuell oder gemäß einem durch Schutzrichtlinien vorgegebenen Zeitplan auf Verzeichnisebene aufgenommen werden. Um Snapshots automatisch zu löschen, können Snapshots mithilfe von Schutzrichtlinien so konfiguriert werden, dass sie für einen bestimmten Zeitraum aufbewahrt werden. Ebenso können Snapshots verschachtelt werden. Snapshots können als unveränderlich definiert werden, um ihre Löschung zu verhindern. Unveränderliche Snapshots können nur dann gelöscht oder geändert werden oder einen kürzeren Aufbewahrungszeitraum erhalten, wenn der HPE Support beauftragt und die Identität des Antragstellers verifiziert wird.

Replikation

Es kann eine Replikation von einem HPE GreenLake for File Storage System zum anderen konfiguriert werden. Der Schutz durch native Peer-Replikation kann pro Verzeichnis konfiguriert und in Snapshots integriert werden. Beim ersten Release wird die asynchrone Replikation mit einem RPO zwischen 15 Sekunden und einigen Minuten unterstützt.

Der Workflow für den Schutz einer Dateifreigabe durch Peer-Replikation gestaltet sich wie folgt:

1. Richten Sie zwei Systeme mit einer Peer-to-Peer-Replikationsbeziehung ein und weisen Sie während des Prozesses optional einen speziellen VIP-Pool für Replikationsverbindungen zu.
2. Erstellen Sie eine Schutzrichtlinie zur Definition der Attribute eines Replikationssystems, beispielsweise RPO, Startzeitpunkt der Replikation und Aufbewahrungszeiträume der lokalen und replizierten Snapshots.
3. Weisen Sie die Schutzrichtlinie einem beliebigen Verzeichnispfad zu, um die Daten durch Replikation zu schützen.

Je nach VIP-Poolzuweisung kann die Replikation hochgradig parallelisiert werden. Allerdings werden Daten, die über die Leitung repliziert werden, komprimiert und nicht durch ähnlichkeitsbasierte Reduktion oder globale Deduplizierung dedupliziert. In anderen Worten: Wenn zwei Exemplare derselben Datei auf dem Quellsystem

vorhanden sind, wird bei der Replikation dieselbe Datei zweimal gesendet. Anschließend wird sie im Zielsystem weiter dedupliziert.

Nach dem Beginn des Replikationsprozesses werden die replizierten Daten unter dem angegebenen Pfad auf dem Zielsystem nicht automatisch mit einer Freigabeeinstellung versehen. Daher sind die replizierten Daten noch nicht als mountfähige Dateifreigabe verfügbar. Um sie einem beliebigen Client als mountfähige Dateifreigabe zur Verfügung zu stellen, muss dem angegebenen Pfad auf dem Zielsystem manuell eine Freigabeeinstellung zugewiesen werden. Die Ziel-Dateifreigabe auf dem Zielsystem hat einen schreibgeschützten Zugriff auf alle Clients. Die Quell-Dateifreigabe behält hingegen den Lese- und Schreibzugriff auf Clients auf dem Quellsystem, wie in Abbildung 7 illustriert. Die bidirektionale Replikation wird unterstützt, solange sich die direkten Pfade beider Replikationsrichtungen nicht überlagern.

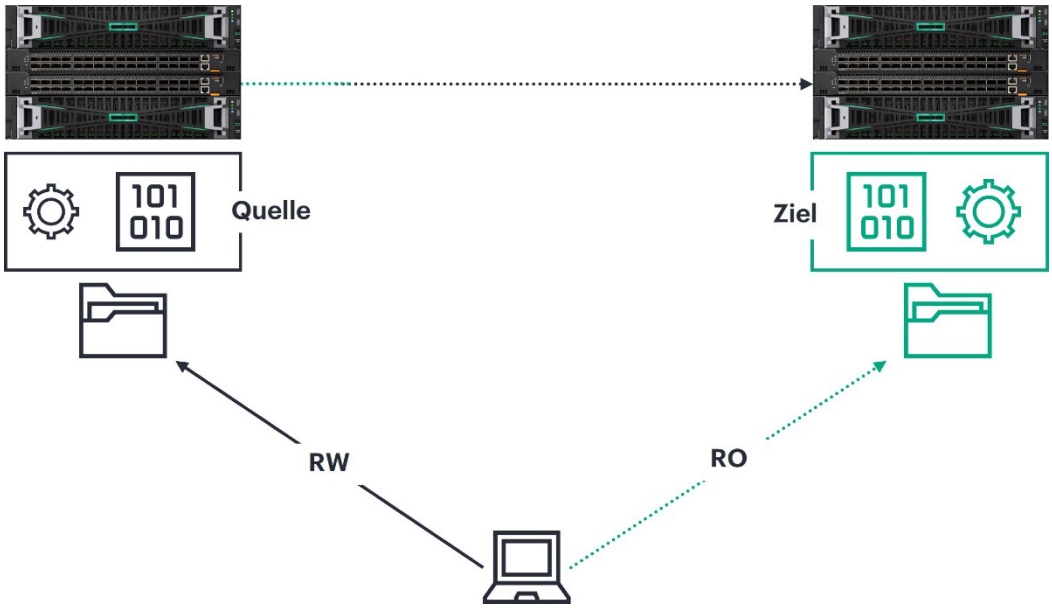


Abbildung 7. Startstatus eines Peer-Replikationssystems

Es sind zwei Arten von Failover verfügbar: **Graceful** und **Ungraceful**. Ein Graceful-Failover kann für jeden geschützten Pfad nur vom Zielsystem initialisiert werden. Damit wird Ihnen ein Failover einer Dateifreigabe ohne Failover eines ganzen Systems ermöglicht. Wird ein Graceful-Failover ausgelöst, führt das System den folgenden Prozess aus:

1. Die Quell-Dateifreigabe wird schreibgeschützt.
2. Ein Snapshot wird im Quellverzeichnis aufgenommen.
3. Die zuvor schreibgeschützte Ziel-Dateifreigabe auf dem Zielsystem kann mit den aktuellen Daten gelesen und beschrieben werden.
4. Die Replikationsrichtung wird umgekehrt und künftige Snapshots werden von der ehemaligen Ziel-Dateifreigabe aufgenommen.

Abbildung 8 zeigt mit der resultierenden Replikationsrichtung und dem Client-Zugriff, was nach dem Abschluss eines Graceful-Failover-Prozesses passiert. Ein Graceful-Failback kehrt den Graceful-Failover-Prozess um. Die daraus resultierenden Status sind die gleichen wie vor dem Failover.

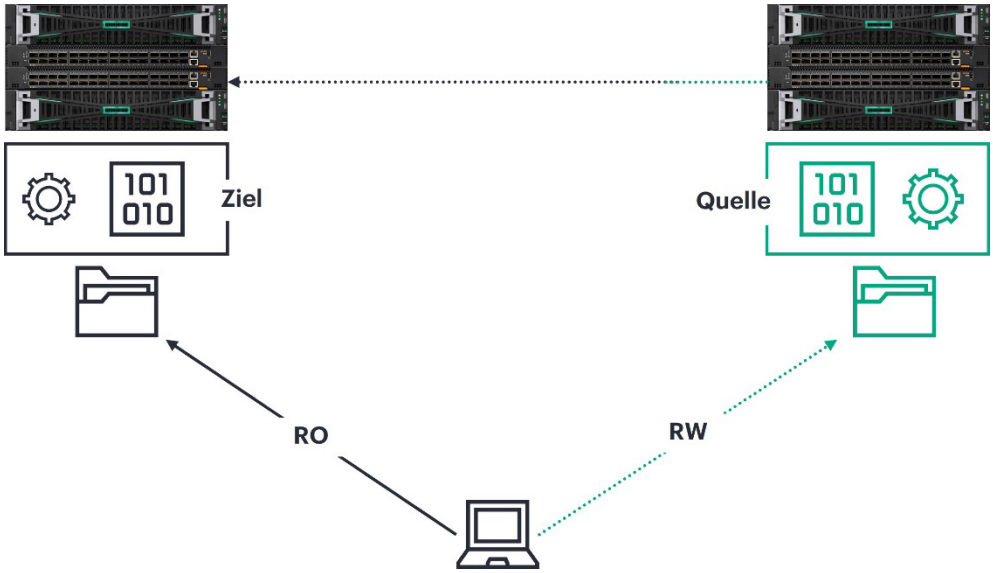


Abbildung 8. Peer-Replikationsstatus nach dem Failover

Ein Ungraceful-Failover macht die Ziel-Dateifreigabe auf dem Zielsystem sofort beschreibbar und beendet die Replikation. Clients erhalten nur dann Schreibzugriff auf die replizierten Daten, wenn die Freigabeeinstellung konfiguriert und dem replizierten Verzeichnis zugewiesen wurde. Snapshots werden unabhängig voneinander lokal auf dem Quell- und Zielsystem aufgenommen. Ein Failback nach einem Ungraceful-Failover erfordert zusätzliche Aufmerksamkeit und Vorbereitung, da während des Failovers beide Exemplare der Dateifreigabe auf dem Quell- und Zielsystem den Schreibzugriff auf Clients beibehalten. In diesem Fall kann sich der Inhalt des replizierten Exemplars auf dem Zielsystem vom Inhalt auf dem Quellsystem unterscheiden. Nach dem Failback wird die Dateifreigabe auf dem Zielsystem schreibgeschützt und verliert die Deltainformationen zu den Änderungen, die während des Ungraceful-Failovers aufgetreten sind. Allerdings bleiben die abweichenden Inhalte in Form von Snapshots auf dem Zielsystem erhalten.

Globale Snapshot-Klone

Globale Snapshot-Klone bieten Clients sofortigen Schreibzugriff auf einen Snapshot der Daten auf einem Remote-Standort, normalerweise 10 Sekunden oder weniger nach der Erstellung des Globalen Snapshot-Klons (Abbildung 9). Der Klon kann aus einem Snapshot entweder auf einem Replikations-Peer-Cluster oder auf dem lokalen Cluster erstellt werden. Globale Snapshot-Klone können für die Replikation in zwei Modi konfiguriert werden. Der erste ist die aktivierte Hintergrundsynchronisierung. In diesem Fall werden die Snapshot-Daten direkt nach der Erstellung des Klons von der Quelle zum Ziel kopiert. Während der Kopierphase werden Lese-Anfragen an die Quelle weitergeleitet, wenn die angeforderten Daten am Remote-Standort noch nicht verfügbar sind. Der zweite Modus ist die deaktivierte Hintergrundsynchronisierung. Dies bedeutet, dass die Snapshot-Daten nicht an das Ziel kopiert werden, außer wenn eine Lese-Anfrage für die Daten vorliegt.

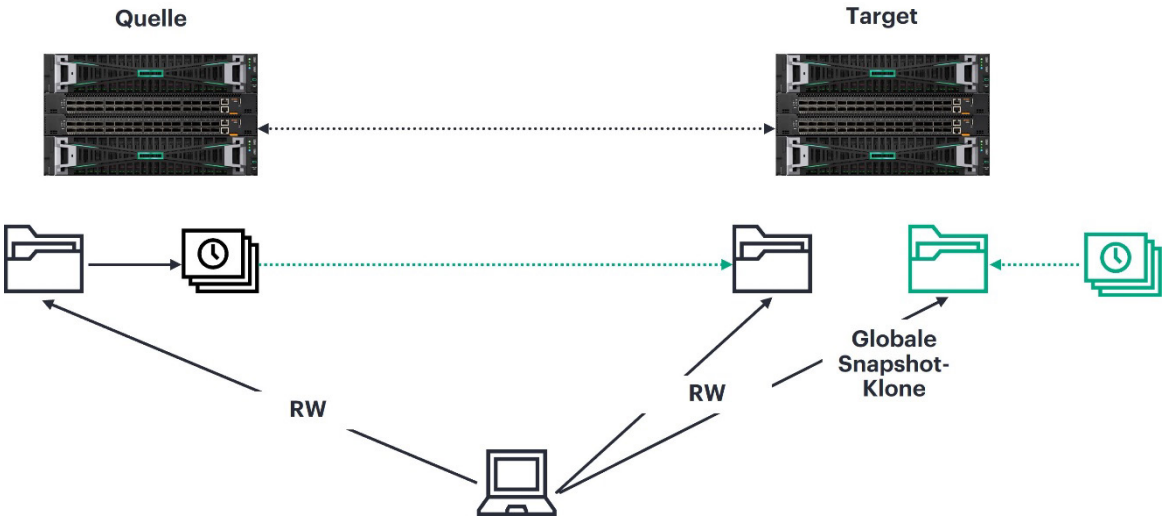


Abbildung 9. Diagramm globaler Snapshot-Klone

Gruppenreplikation

Die Gruppenreplikation ermöglicht eine 1:N-Replikation auf demselben Datenpfad mit unterschiedlichen Richtlinien und Zeitplänen für mehrere Remote-Standorte. Die Daten werden während eines Failovers mithilfe gemeinsamer Resynchronisierungspunkte intelligent neu synchronisiert. Dies ermöglicht nach einem Failover einen deutlich schnelleren Zugriff auf die neueste Version der Daten am Remote-Standort. Resynchronisierungspunkte können explizit von Benutzern oder implizit durch überlappende Zeitpläne definiert werden.

Ein typischer Datenpfad einer konfigurierten Gruppenreplikation könnte folgendermaßen aussehen:

1. Eine Gruppenreplikation ist so konfiguriert, dass sie vom Quellcluster A auf zwei Remote-Standorte B und C repliziert. Die Daten werden stündlich auf B repliziert, während die gleichen Daten auch alle 30 Minuten auf C repliziert werden.
2. Bei einem Failover von Cluster A wird Cluster B zum neuen Quellcluster und die Daten können somit durch Lese- und Schreibvorgänge abgerufen werden. Während des Failovers von Cluster A müssen die Daten auf Cluster B nur eine teilweise Neusynchronisierung mit Cluster C für die Deltas zwischen den neuesten verfügbaren Kopien auf beiden Clustern durchführen. Die Replikationen von Cluster B werden weiterhin von B nach C und nach A repliziert, wenn A erreichbar ist.
3. Wenn der Failover dazu dient, den Remote-Standort in den Standalone-Modus zu versetzen, wird Cluster B nach dem Failover nicht mehr nach C oder A repliziert.

Sicherheit

HPE GreenLake for File Storage bietet eine Vielzahl von Sicherheitsfunktionen für einen rundum besseren Schutz, darunter Authentifizierung, Verschlüsselung, Auditierung und Mandantenfähigkeit für einen isolierten Datenzugriff sowie die rigorosen Sicherheitsmerkmale der verwaltungsfreundlichen HPE GreenLake Plattform.

Datenberechtigung und -authentifizierung

Sämtlicher Datenzugriff wird über eine interne Benutzerdatenbank kontrolliert, gepflegt und gespeichert. Zugriffsbits werden als POSIX-Modus-Bits, POSIX-ACLs, SMB-ACLs und NFSv4-ACLs dargestellt. Alle externen Anbieter und lokalen Benutzer und Gruppen, beispielsweise UID, GID, SID usw., werden mit internen IDs in der internen Benutzerdatenbank verknüpft. Die interne Benutzerdatenbank dient als Zwischenspeicher für jeden Benutzer und wird über das Managementnetzwerk regelmäßig durch die konfigurierten externen Authentifizierungsanbieter, beispielsweise AD, NIS oder LDAP, aktualisiert. Die interne Benutzerdatenbank befindet sich auf dem SCM-Layer und wird auf alle verfügbaren SCM-Geräte gespiegelt und verteilt. Bedenken Sie, dass die

Skalierbarkeit des Zwischenspeichers von zahlreichen Benutzereinträgen von externen Anbietern durch die SCM-Kapazität begrenzt wird.

Verschlüsselung ruhender Daten

Sämtliche Benutzer- und Metadaten auf einem HPE GreenLake for File Storage System können im Ruhezustand optional verschlüsselt werden. Diese Option kann während der Installation konfiguriert werden. Die Verschlüsselungsschlüssel werden intern aufbewahrt. Die Funktion zur Verschlüsselung ruhender Daten funktioniert unabhängig von allen Dateiprotokoll-Clients.

Auditierung

Die Auditierung kann für Protokollmerkmale wie NFS oder SMB sowie für Arten von zu auditierenden Dateivorgängen, beispielsweise Erstellen, Löschen usw., konfiguriert werden. Ebenso ist eine Auditfunktion für Managementdienst-Aktivitäten verfügbar und überwacht Managementaktivitäten im gesamten System. Die Kapazität der Auditierung kann pro CNode eingeschränkt werden. Die Aufbewahrung kann entweder für einen festen Zeitraum oder so konfiguriert werden, dass die Auditdateien für immer erhalten bleiben. Die Auswirkungen auf Leistung und Kapazität variieren je nach dem konfigurierten Auditierungsniveau.

Mandantenfähigkeit

Die Mandantenfunktion von HPE GreenLake for File Storage bietet zusätzliche Optionen für die Datenisolierung und Datendomänen. Weitere Informationen finden Sie im Abschnitt Controller-Pooling, QoS und Mandantenfähigkeit dieses Whitepapers.

Sicheres Cloud-basiertes Management

Das Cloud-basierte Management von HPE bietet zahlreiche Vorteile sowohl für die Dateninfrastruktur als auch die Daten selbst. Die Data Services Cloud Console ist die Cloud-basierte Anwendung von HPE für das Daten- und Speichermanagement heute und in Zukunft. Sicherheit steht bei der Entwicklung der Data Services Cloud Console im Mittelpunkt. Bei ihrer Entwicklung und der Entwicklung der zugehörigen Produkte hat HPE speziell an die Sicherheitsanforderungen seiner Kunden gedacht. Weitere Informationen zu sämtlichen Sicherheitsaspekten von HPE GreenLake sowie der Data Services Cloud Console finden Sie im Whitepaper „Sicherheitsleitfaden für Data Services Cloud Console“.

Sonstige Sicherheitsfunktionen

In den letzten Updates von v3.1 wurden einige Sicherheitsfunktionen eingeführt.

Die erste ist das NFSv4.2-Sicherheitsetikett, ein Umschalter, der für jeden Mandanten aktiviert oder deaktiviert werden kann. Die Sicherheitsetikette von NFSv4.2 sind für die Zusammenarbeit mit SELinux konzipiert, um obligatorische Zugriffskontrollen bereitzustellen, die als erweiterte Attribute gespeichert werden. Diese Funktion ist nur anwendbar, wenn SELinux-Sicherheitsetikette erforderlich sind.

Die zweite in v3.1 eingeführte Sicherheitsfunktion ist die attributbasierte Zugriffskontrolle (ABAC) als Teil der Zero-Trust-Funktionen. Die ABAC-Funktion realisiert eine Zugriffskontrolle basierend auf Attributen, die mit Dateien oder Ordnern verknüpft sind, und trennt die Zugriffskontrolle vom Speicheradministrator. Sie verlässt sich stattdessen auf den Authentifizierungsanbieter. ABAC-Tags können pro Dateifreigabe ausgewählt werden, sie werden jedoch zum Zeitpunkt der Erstellung nicht gegenüber dem Authentifizierungsanbieter ausgewertet und können nach der Erstellung nicht mehr geändert werden. Benutzer mit ABAC-Tags werden viel häufiger aktualisiert, alle 60 Sekunden oder manuell. ABAC-Tags werden vor und in Verbindung mit den Dateisystemberechtigungen und ACLs ausgewertet, und die restriktivste Berechtigungskombination hat Vorrang. Bei Vorhandensein von mehreren ABAC-Tags ist ebenfalls die restriktivste Kombination ausschlaggebend.

Die dritte Funktion ist die WORM-Funktion (Write-Once, Read-Many), die für NFSv3- und SMB-Dateifreigaben verfügbar ist. Es gibt zwei Möglichkeiten, die Aufbewahrungsfrist für Dateien für WORM zu definieren, da es bei Datei-Workloads oft schwierig ist, zu bestimmen, wann ein Benutzer mit der Bearbeitung einer Datei fertig ist. Die erste Möglichkeit, den Übergang einer Datei oder eines Ordners in den schreibgeschützten Modus zu konfigurieren, besteht darin, als Zeitpunkt einen zukünftigen Zeitpunkt zuzuweisen. Das bedeutet, dass die Datei

beschreibbar bleibt, solange das Aufbewahrungsdatum nicht erreicht ist. Die zweite Möglichkeit ist das Auto-Commit zum WORM-Modus, der eine Aufbewahrungszeit auf Dateifreigabeebene festlegt, sodass die Datei(en) und Ordner schreibgeschützt werden, sobald die automatisch berechnete Zeit seit der Erstellung abgelaufen ist.

Speichermanagement

Die Benutzeroberflächen von HPE GreenLake for File Storage erleichtern Datenspeicher-Admins die Arbeit beim Management von Datei-Workloads. Die Data Services Cloud Console ermöglicht ein einfaches Datamanagement mit einem intuitiven Cloud-Nutzungserlebnis sowie alle nötigen Funktionen für das Management Ihrer HPE Infrastruktur und Workloads (z. B. Block, Datei, Sicherung und Wiederherstellung) – alles an einem Ort. Die lokale Administrations-UI steht auch für das Management einzelner Arrays über eine GUI zur Verfügung.

Einige der in der Onboard UI und in der Data Services Cloud Console verwendeten Begriffe unterscheiden sich möglicherweise von den gebräuchlichen NAS-Begriffen. Diese Begriffe werden in Tabelle 4 im Anhang zusammengefasst.

Data Services Cloud Console

Die Data Services Cloud Console ist eine Anwendung, die auf HPE GreenLake für alle Ihre Anforderungen rund um Data Services verfügbar ist. Durch die Beseitigung von isoliertem Infrastrukturmanagement macht sie das Management weniger komplex. Die Data Services Cloud Console ist eine in der Cloud verwaltete zentrale Konsole, die von jedem Ort und Gerät aus erreicht werden kann. In der Data Services Cloud Console finden Sie zwei Services für das Management Ihres Dateispeichers:

- **Data Ops Manager** ermöglicht Ihnen das Management der verschiedenen Datenspeicherplattformen Ihres Unternehmens, unter anderem HPE Alletra 6000, HPE Alletra 9000 und HPE Alletra Storage MP. Sie können Data Ops Manager zur Anzeige Ihres Systembestands und der Systemdetails nutzen. In der Systemliste können Sie Systeme filtern, durchsuchen und sortieren. Dabei haben Sie Zugriff auf Datenspeicher-Arrays für Dateien und Blöcke. Im Dashboard sehen Sie den allgemeinen Array-Status, die Kapazität und die Leistung. Neben den allgemeinen Array-Informationen finden Sie die nützlichsten Details in den systemspezifischen Reitern. Beispielsweise sehen Sie im Fenster „Issues" (Probleme) alle kritischen Probleme, die nach einem Eingriff des Datenspeicher-Administrators verlangen. Die Kapazität und Leistung des Systems ist einfach zu erkennen, wenn Sie die historischen Metriken betrachten. Darüber hinaus können Datenspeicher-Administratoren über Data Ops Manager Konfigurationen auf Systemebene ändern, indem sie beispielsweise VIP-Pools, Benutzer und Gruppen, Replikation, Netzwerkeinstellungen usw. erstellen oder verändern.
- **Das Dashboard der File Storage Anwendung** bietet eine einheitliche aggregierte Ansicht aller Ihrer HPE GreenLake for File Storage Systeme, einschließlich sämtlicher Leistungs-, Kapazitäts- und sonstiger wichtiger Metriken und Details zur Datenspeichernutzung. Jede Kachel kennzeichnet die wichtigsten Dateifreigaben oder Clients mit potenziellen Hot Spots. In der Liste der Dateifreigaben sehen Sie Freigaben auf Basis von Merkmalen wie Protokollen, Schutzrichtlinien und Datenspeicher-Arrays. Alternativ können Sie sie auch in einer leistungsorientierten Ansicht anzeigen lassen. In dieser Ansicht können Sie Dateifreigaben nach Nutzung, Anzahl an Dateien, Kontingent und Leistungsmetriken anzeigen und filtern. In den Details jeder Dateifreigabe finden Sie Angaben zu Kapazität, Leistung, Benutzer- und Gruppen-Accounts sowie zum Schutz.

In der File Storage Anwendung können Sie Dateifreigaben sowie Einstellungen für Dateifreigaben erstellen. In den allgemeinen Einstellungen können Sie Parameter wie beispielsweise den Fileserver definieren, der als Host für die Freigaben dient, auf die die Einstellungen angewendet werden:

- **Der Sicherheitsfaktor** legt die Protokollregeln fest, denen die Dateivorgänge folgen werden.
- **VIP-Pools** dienen dem Lastausgleich und der QoS.
- **Authentifizierungsquellen** können als lokale oder externe Anbieter gekennzeichnet werden.

– **Mindestschutzniveaus für den Hostzugriff** für NFS 4.1 können auf eine Auswahl von Client-Systemen angewendet werden.

– **Kerberos-Anmeldeinformationen** können für zusätzliche Sicherheit genutzt werden.

Für eine noch stärkere Sicherheitskontrolle können Sie in den hostbasierten Zugriffseinstellungen Gruppen oder einzelne Clients für protokollspezifische Zugriffsregeln festlegen, beispielsweise Lese- und Schreibzugriff und Root-Squash auf Client-Basis.

Onboard UI

Onboard UI ist eine webbasierte GUI für die Administration und Überwachung des HPE GreenLake for File Storage Systems. Sie steht Benutzern mit einer Managementnetzwerk-Verbindung zu den CNodes per HTTP zur Verfügung. Jede über Onboard UI verwaltete Instanz wird auf dem Cluster gehostet und verwaltet diesen Cluster. Onboard UI bietet zahlreiche Optionen für Konfigurationen, Datenspeicherverbrauch sowie Echtzeit- und historische Analysen.

HPE iLO

HPE iLO ist eine webbasierte GUI, die eine sichere Konfiguration, Überwachung und Aktualisierung von Hardwarekomponenten ermöglicht, und ist auf jedem CNode und DNode installiert. Es ist der Schlüssel, um die CNodes und DNodes betriebsfähig zu machen und zu starten. Es vereinfacht die Einrichtung und Aktivierung der Zustandsüberwachung sowie der Energie- und Wärmesteuerung. Diese branchenführenden Funktionen erhöhen die Produktivität von Serveradministratoren. Weitere Informationen zu HPE iLO finden Sie auf der Produktseite von HPE iLO Servermanagement.

API

Die HPE GreenLake for File Storage API bietet Ihnen ebenfalls eine leistungsstarke und flexible Möglichkeit, Ihren HPE Dateispeicher zu verwalten. Diese API stellt Ihnen Funktionen für das programmatische und autonome Management ihres gesamten HPE GreenLake Dateispeichers zur Verfügung. Mit ihrer Hilfe lässt sich das Dateispeicher-Management mühelos in Ihren Automatisierungsprozess integrieren. Die HPE GreenLake for File Storage API ermöglicht das Management von Dateifreigaben und Einstellungen sowie Hardware über eine Reihe von HTTPS-Anfragen. Die API umfasst einen Server, der Teil von HPE GreenLake for File Storage auf Basis des HPE Alletra MP Betriebssystems ist und sowohl auf dem Datenspeichersystem als solches als auch auf einer Definition der Vorgänge, Eingaben und Ausgaben der API ausgeführt wird.

Unterstützung von Container-Workloads

Container werden immer häufiger für Produktionsanwendungen in Unternehmen eingesetzt, da immer mehr Unternehmen von der Flexibilität und Effizienz von Containern profitieren. HPE GreenLake for File Storage stellt nicht nur die standardmäßige NFS-Freigabe für das Mounten von Pods bereit, sondern unterstützt auch das Container Storage Interface (CSI) von VAST Data. Dieser CSI-Treiber bietet Funktionen für dynamische Bereitstellung, Volume-Erweiterung, mehrere Datenspeicherklassen, Snapshots von persistenten Volumes und flüchtige Volumes. Da die Architektur in der Lage ist, sämtlichen verfügbaren Datenspeicher für einen einzelnen Mountpunkt offenzulegen, hängt die Kapazität eines dynamisch bereitgestellten Volumes von dem Kontingent ab, das automatisch im Datenspeicher-Array festgelegt und erweitert werden kann.

Zusammenfassung

Unternehmen wachsen immer weiter und erzeugen immer mehr unstrukturierte Daten in einem beispiellosen Tempo. Daher ist es entscheidend, zukunftssichere Lösungen zu implementieren, die nicht nur mit den steigenden Anforderungen des Zeitalters von künstlicher Intelligenz und maschinellem Lernen Schritt halten, sondern auch Einblicke in die erzeugten Daten liefern können, um für bessere Geschäftsergebnisse zu sorgen. Mithilfe der in diesem Dokument geschilderten architektonischen Grundlagen können Sie die HPE GreenLake for File Storage Lösung effektiv zusammenstellen und implementieren, um Ihre geschäftlichen Herausforderungen auf dem Markt für unstrukturierte Daten zu bewältigen.

Anhang

Tabelle 4. Begriffe im Zusammenhang mit HPE GreenLake for File Storage

Häufig verwendete Begriffe	HPE GreenLake for File Storage	Definitionen
Gehäuse	CBox	Computing-Node-Gehäuse mit je zwei Controller-Nodes
Computing-Node	CNode	Computing-Nodes oder Computing-Nodes-Eingabe-/Ausgabemodule (IOMs)
JBOF-Gehäuse	DBox	JBOF-Chassis, in dem sich die SCMs und SSDs befinden, die jeweils mehrere JBOF-IOMs enthalten können
JBOF-Node	DNode	JBOF IOM
Laufwerke	SSD, NVRAM	Festplattenlaufwerke, einschließlich NVMe-SSD-Laufwerken und NVMe-SCM- oder NVRAM-Laufwerken
Systemeinstellungen	Cluster-Einstellungen	Systemweite Konfigurationen
Dateifreigabe	Dateifreigabe, Ansicht	Ein Verzeichnispfad, dem Dateifreigabe-Einstellungen/Ansichtsrichtlinien zugewiesen sind und der zu einer mountfähigen Dateifreigabe für einen Client wird
Freigabe-Einstellungen	Freigabe-Einstellungen, Ansichtsrichtlinien	Ein Satz von Attributen zur Definition der grundlegenden Attribute zur Bestimmung des dateiprotokollspezifischen Verhaltens
Lokale Benutzer/Gruppen	Benutzer/Gruppen	Benutzer und Gruppen, die im HPE GreenLake for File Storage Cluster definiert sind; davon ausgenommen sind Benutzer und Gruppen von externen Authentifizierungsanbietern

Hinweis
HPE hat keine Kontrolle über Websites Dritter und übernimmt keine Verantwortung für diese (oder dort aufgeführte Produkte, Dienstleistungen oder Inhalte).

Ressourcen

HPE GreenLake for File Storage

HPE GreenLake for File Storage – Technische Demo

Sicherheitsleitfaden für Data Services Cloud Console – Whitepaper

HPE iLO Server Management-Produktseite

Weitere Informationen

HPE GreenLake for File Storage

HPE.com besuchen

Jetzt chatten

© Copyright 2025 Hewlett Packard Enterprise Development LP. Die hier enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden. Neben der gesetzlichen Gewährleistung gilt für Produkte und Services von Hewlett Packard Enterprise (HPE) ausschließlich die Herstellergarantie, die in den Garantieerklärungen für die jeweiligen Produkte und Services explizit genannt wird. Die hier enthaltenen Informationen stellen keine zusätzliche Garantie dar. Hewlett Packard Enterprise haftet nicht für hierin enthaltene technische oder redaktionelle Fehler oder Auslassungen in diesem Dokument.

AMD ist eine Marke von Advanced Micro Devices, Inc. Docker ist eine Marke oder eingetragene Marke von Docker, Inc. in den USA und/oder anderen Ländern. Intel ist eine Handelsmarke der Intel Corporation oder ihrer Tochtergesellschaften in den USA und/oder anderen Ländern. Linux ist in den USA und anderen Ländern eine eingetragene Marke von Linus Torvalds. Active Directory ist in den USA und/oder anderen Ländern eine eingetragene Marke oder Marke der Microsoft Corporation. GPUDirect, NVIDIA DGX, das NVIDIA-Logo und NVIDIA sind Marken und/oder eingetragene Marken der NVIDIA Corporation in den USA und anderen Ländern. Alle Marken Dritter sind Eigentum ihrer jeweiligen Inhaber.

a00132707DEE, Rev. 4

HEWLETT PACKARD ENTERPRISE

hpe.com