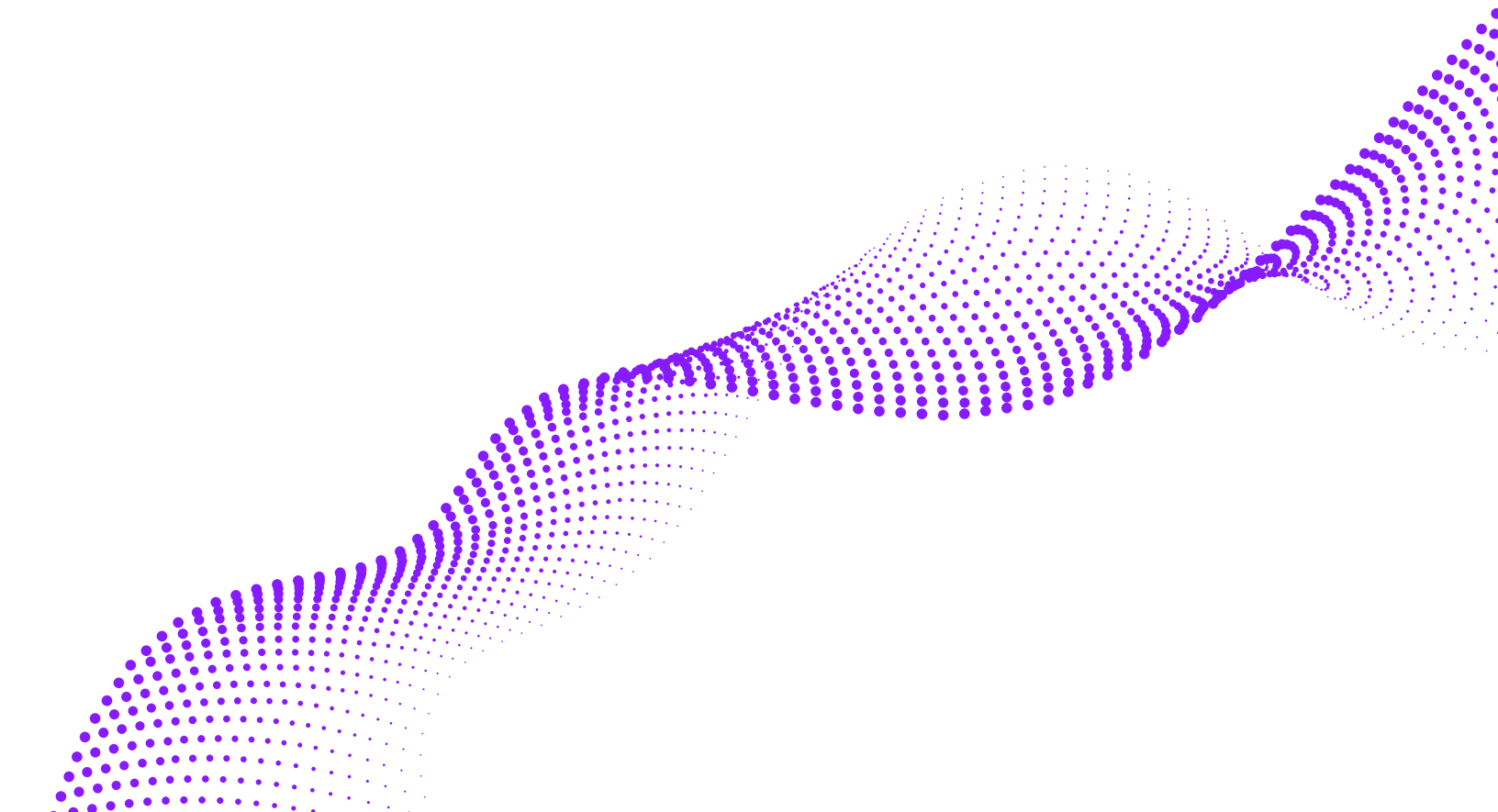
Lenovo
INFINIDAT

WHITE PAPER

Infinidat-Speicherarchitektur

Wettbewerbsvorteile im Petabyte-Bereich



Zusammenfassung

Infinidat®-Speicherlösungen für Unternehmen basieren auf der einzigartigen und patentierten Infinidat-Speicherarchitektur – einem vollständig abstrahierten Satz von Software-Defined Storage (SDS)-Funktionen, integriert in die beste Standard-Hardware. Die softwareorientierte Architektur von Infinidat steht für eine Evolution und Revolution im Datenmanagement der letzten 30 Jahre und löst die widersprüchlichen Anforderungen an größere, schnellere und gleichzeitig kostengünstigere Speicher. In diesem Dokument wird die Technologie beschrieben, mit der Infinidat als einziger Anbieter von Unternehmensspeichern eine Kapazität von mehreren Petabyte mit bahnbrechender Leistung (35 Mikrosekunden Latenzzeit1) und einer beispiellosen 100%igen Verfügbarkeit erreicht, und das alles bei niedrigsten Gesamtbetriebskosten (TCO).

[1]Interne Latenz von nur 35 Mikrosekunden, basierend auf Telemetriedaten aus kundenseitigem Produktivbetrieb, Lese-E/A-Treffer (DRAM-Zwischenspeicher). Die tatsächlichen Ergebnisse können abweichen. März 2022.

Designprinzipien

Beim Entwurf einer Speicherarchitektur, die den Anforderungen moderner Rechenzentren entspricht, müssen mehrere Anforderungen erfüllt werden:

KATEGORIE	ANFORDERUNG
Zuverlässigkeit	Unternehmen arbeiten rund um die Uhr. Ausfallzeiten werden nicht toleriert.
Kapazität	Exponentiell wachsende Datenmengen, beschleunigt durch die digitale Transformation, heterogene Big-Data-Architekturen, künstliche Intelligenz (KI) und maschinelles Lernen (ML)
Leistung	Die Leistung muss mit der Skalierung der Daten muss Schritt halten, um die gleichen (oder bessere) Ergebnisse in den gleichen (oder kürzeren) Zeiträumen zu erzielen.
Einfachheit	Administratoren erwarten einfache Abläufe, eine umfassende Integration des Ökosystems und integrierte Tools für den Übergang zu DevOps-Modellen, um weniger Zeit mit der Speicherverwaltung zu verbringen und mehr Zeit für Anwendungen und Geschäftsprozesse zu haben.
Konsolidierung	Point-Technologien gehören der Vergangenheit an: Moderne Speicher sollten alle Anwendungsfälle abdecken, um maximale Effizienz, Einfachheit und Kosteneinsparungen zu erzielen.
Kosten	Die Budgets passen sich nicht an die Kapazitäts- und Leistungssteigerungsanforderungen an. Eine radikale Änderung der Architektur ist erforderlich.
Resilienz	Unternehmen benötigen Widerstandsfähigkeit sowohl innerhalb ihrer Infrastruktur als auch gegenüber externen Bedrohungen wie Cyberangriffen.

Gleichzeitig behaupten Anbieter von öffentlichen Cloud-Infrastrukturen wie Amazon, Google und Azure, die Kosten für den gesamten IT-Stack zu senken. Das tun sie auch oft für kleine Kunden, die sich kein großes IT-Personal leisten können und die sich oft auf ein oder zwei „Alleskönner“ verlassen, um ihren gesamten IT-Betrieb aufrechtzuerhalten. Für große Unternehmen sowie regionale Cloud- und Managed-Services-Anbieter ist es jedoch wichtig, einen effizienteren IT-Stack einzuführen. Dieser muss ihren geschäftlichen, technologischen und finanziellen Anforderungen entsprechen und ihnen alle Vorteile der Cloud innerhalb der eigenen Infrastruktur bieten, während sie gleichzeitig die Kosten senken und die Hoheit über ihre Daten behalten.

InfiniBox®-Architektur

Infinidats Vorführprodukt InfiniBox® wurde unter Berücksichtigung der wichtigsten Prinzipien entwickelt, um allen diesen Herausforderungen gerecht zu werden:

PRINZIP	BEGRÜNDUNG	HERAUSFORDERUNG
Innovatives Softwaredesign	Im Gegensatz zu Hardware wird Software im Laufe der Zeit optimiert und verbessert die Leistung, anstatt sie zu verschlechtern. InfiniBox basiert auf über 140 Softwarepatenten. Das ist die wahre Bedeutung von Software Defined Storage (SDS). Die innovative Software von Infinidat umfasst unter anderem die patentierten Neural Cache-Algorithmen, Metadaten-Management-Strukturen und Speicherfunktionen der nächsten Generation.	**Leistung** **Einfachheit** **Zuverlässigkeit** **Kosten**
Auf Resilienz ausgelegt	Beim Entwurf einer skalierbaren Lösung ist die Resilienz entscheidend. InfiniBox ist auf sieben Neunen (99,99999 %) Zuverlässigkeit ausgelegt und verwendet eine dreifach redundante Architektur, bei der alle kritischen Komponenten (Software und Hardware) über mindestens zwei (2) Redundanzen (N+2) verfügen, die vor Ausfallzeiten und Datenverlust schützen. Darüber hinaus bietet Infinidat eine 100%ige Systemverfügbarkeitsgarantie für InfiniBox und InfiniBox SSA, die auf Zehntausenden von Betriebsstunden an Erfahrung basiert und durch unabhängige und branchenführende Gartner Peer Insights Reviews unterstützt wird.	**Resilienz** **Kosten** **Einfachheit** **Konsolidierung**
Skalierbare Architektur	Um Kapazität und Leistung zu disruptiven Kosten zu erreichen, bedarf es einer Skalierung. InfiniBox wurde für Großkunden entwickelt, die eine Skalierung auf 10 PB oder mehr effektive Kapazität in einem einzigen 42U-Rack benötigen.	**Konsolidierung** **Kosten** **Einfachheit** **Kapazität**
Ausfallsichere Integration von Hardware und Software	Infinidat evaluiert alle führenden Anbieter von Hardwarekomponenten und wählt nur die zuverlässigsten aus, um sie im Rahmen der InfiniBox-Lösung einzusetzen. Dank diesem Best-of-Breed-Ansatz erhalten die Kunden eine vollständig integrierte und getestete Lösung – im Gegensatz zur typischen SDS, die eine komplizierte Hardwareintegration und Administration erfordert.	**Zuverlässigkeit** **Einfachheit** **Konsolidierung**
Standardhardware (COTS)	Die Verwendung von Standardhardware und die Vermeidung langer Entwicklungszyklen vereinfachen die Einführung neuer Technologien. Dazu gehören CPUs, Speichertypen und Speichermedien. Die Verwendung von Standardhardware und der dazugehörigen Software bewirkt auch mehr Stabilität, da die gleiche Hardware in Tausenden von Systemen weltweit eingesetzt wird.	**Kosten** **Zuverlässigkeit** **Kapazität** **Einfachheit** **Leistung**

Aufbau auf starker Architektur

Die InfiniBox-Architektur bietet auch weiterhin beispiellose Leistung, Zuverlässigkeit, Skalierbarkeit, Einfachheit und Wirtschaftlichkeit. Diese mit InfiniBox eingeführte Architektur bildet auch die Grundlage für weitere Innovationen. Heutzutage bietet Infinidat noch mehr Auswahl, Leistung und Skalierbarkeit, was die Vorzüge dieser gemeinsamen Architektur unterstreicht.

InfiniBox – Speicher der Unternehmensklasse mit KI-basiertem Neural Cache.

InfiniBox SSA – Alle Solid-State-Komponenten für die anspruchsvollsten Workloads.

InfiniGuard – Speziell entwickelte Datenschutz- und Wiederherstellungs-Appliance mit integrierter Cyber-Resilienz.

Die Stärke dieser Architektur hat sich in den letzten zehn Jahren durch die konsequente Bereitstellung von Innovationen, Wettbewerbsvorteilen und Kundennutzen gezeigt. Und das wird auch in Zukunft so sein.

Leistungsbeschleunigung

InfiniBox verwendet eine Kombination aus DRAM, Flash-Medien (SSD) und NL-SAS-Festplatten mit hoher Kapazität zum Schreiben, Lesen und Speichern von Daten. Im Folgenden wird erläutert, wie Lese- und Schreibvorgänge beschleunigt werden, um maximale Leistung bei minimaler Latenz zu erreichen. Die interne Latenzzeit, die Infinidat-Kunden mit InfiniBox SSA erleben, beträgt beispielsweise 35 Mikrosekunden. Der zur Optimierung der Datenplatzierung verwendete Algorithmus heißt Neural Cache. In diesem Abschnitt wird erläutert, wie Neural Cache den Kunden durch den Einsatz intelligenter Software-Algorithmen die branchenweit niedrigste Latenzzeit bietet. Zu bedenken ist, dass die meisten Transaktionsanwendungen mindestens zwei getrennte I/O erfordern (einen zum Schreiben einer Transaktion in die Protokolle und einen zum Schreiben von Daten in die Datenbank). Damit wird die Latenz zur Schlüsselkomponente bei der Beurteilung sowohl der Anwendererfahrung als auch der maximalen Leistung der Applikation.

METADATENSCHICHT

Die Antwortzeiten der Metadatenschicht wirken sich unmittelbar auf die I/O-Latenz aus. InfiniBox beschleunigt Metadatenoperationen folgendermaßen:

- **Alle Metadaten befinden sich im DRAM**: Die Metadaten werden im DRAM gespeichert, was sowohl Lese- als auch Schrei vorgänge beschleunigt.
- **Metadatenstruktur**: Ein vollständiger Verlauf aller in die InfiniBox geschriebenen Daten wird in einer „Trie“ genannten Metadatenstruktur verwaltet. Diese patentierte Implementierung erfasst die Datenplatzierung, relevante organisatorische und virtuelle Adressierungsinformationen sowie mehrere Ebenen des Datenschutzes.
- **Trie-Effizienz**: Alle Einfügungen, Änderungen und Löschungen im Trie erfolgen mit der gleichen Latenz und bieten eine gleichbleibende Leistung von den ersten Datenbyte bis hin zu mehreren Petabyte.

SCHREIBBESCHLEUNIGUNG

InfiniBox akzeptiert alle Schreibvorgänge ohne Vorverarbeitung (z. B. Musterentfernung, Komprimierung, Verschlüsselung usw.) im DRAM und erstellt eine zweite Kopie des Schreibvorgangs im DRAM eines anderen Knotens über InfiniBand mit niedriger Latenz, bevor die Bestätigung an den Host gesendet wird. Durch die Annahme des Schreibvorgangs vom DRAM (direkt an die CPU angeschlossen) anstelle eines externen Flash-Geräts kann InfiniBox Schreibvorgänge mit der geringstmöglichen Latenzzeit abschließen.

Im Gegensatz zu vielen Architekturen, bei denen der Schreibcache in kleine Bereiche aufgeteilt wird (wie bei Matrix- und Dual-Controller-Architekturen), verwendet InfiniBox einen einzigen, großen Speicherpool für Schreibvorgänge. Dadurch können größere Schreibvorgänge aufrechterhalten und häufig geänderte Daten mit DRAM-Latenz überschrieben werden. Der Neural Cache hat Zeit, intelligente Entscheidungen dahingehend zu treffen, welche Datenblöcke von der DRAM-Geschwindigkeit profitieren und

welche auf SSDs und HDDs ausgelagert werden sollten. Durch die längere Verweildauer der Daten im Schreibcache vermeidet Neural Cache eine unnötige Belastung der CPU und der Back-End-Persistenzschichten.

Vor der Löschung sammelt jeder Cache-Zyklus random geschriebene Daten und fügt sie anhand verschiedener Faktoren wie Relevanz der geschriebenen Daten zu größeren sequential Schreibvorgängen zusammen. Dadurch kann der Neural Cache später bei der prädiktiven Analyse entscheiden, welche Daten in Verbindung mit jedem Lesevorgang benötigt werden.

LESEBESCHLEUNIGUNG

Im Gegensatz zu herkömmlichen Speicher-Arrays, bei denen die aktivsten Daten (auch als „heiße Daten“ oder „Hot Data“ bezeichnet) im Flash-Cache platziert werden, um eine Leistungsparität mit herkömmlichen All-Flash-Arrays zu erreichen, verwendet InfiniBox seinen innovativen Neural Cache, der auf die Platzierung aller heißen Daten im DRAM abzielt. Mit dem InfiniBox Neural Cache können die meisten Lesevorgänge mit DRAM-Geschwindigkeit abgeschlossen werden, was 1000 Mal schneller ist als Flash.

Die globale Data Fabric von Infinidat umfasst viele Exabyte an Daten, und Neural Cache liefert nachweislich fast alle Lesevorgänge aus dem DRAM, so dass Kunden von einem Nutzungskomfort wie mit „All-DRAM-Array“ zu niedrigeren Gesamtbetriebskosten als bei konkurrierenden Arrays profitieren.

Da es sich bei Neural Cache um einen lernenden Algorithmus handelt, wird die Leistung im Laufe der Zeit optimiert. InfiniBox nutzt eine umfangreiche SSD-Flash-Schicht als eine Art Puffer für DRAM-Fehler. Während Neural Cache die I/O-Muster erlernt und die DRAM-Datenplatzierung optimiert, ändert die Flash-Schicht ihre Funktion von der Behandlung von DRAM-cache misses hin zur Behandlung von Änderungen der I/O-Muster, die der Algorithmus möglicherweise nicht vorhersagen kann (z. B. regelmäßige Audits, bei denen nicht im DRAM befindlichen Daten benötigt werden).

Softwarearchitektur

Beim Design von InfiniBox zur Aufrechterhaltung der 100%igen Verfügbarkeit verwendet Infinidat Software, um die Unvorhersehbarkeit von Hardwareausfällen zu meistern. InfiniBox setzt eine Active/Active/Active-Softwarearchitektur und ein laufend überwachwenden, automatisch selbstheilenden und bei Hardwareausfällen auf allen Ebenen zuverlässig wiederherstellenden N+2-Design ein.

Alle Komponenten sind in die Software implementiert, vom RAID bis hin zu den geclusterten Diensten, um eine ständige Optimierung mit jedem neuen Release zu ermöglichen. In den ersten fünf Jahren seit der ersten allgemein verfügbaren Version hat sich die maximale Leistung von InfiniBox um mehr als das 4-Fache verbessert, und das nur durch eine unterbrechungsfreie Aktualisierung der Software. Das ist die Stärke einer echten softwaredefinierten Lösung.

GECLUSTERTE DIENSTE

Alle Datendienste laufen gemäß dem N+2-Architekturdesign auf allen Knoten und sind auf allen Knoten aktiv (keine passiven Knoten im Cluster). Die Datendienste sind so konzipiert, dass sie im Benutzerbereich ausgeführt werden, einschließlich Low-Level-Komponenten wie Fiber Channel (FC)-Treiber. Da im Kernel keine Datendienste laufen, kann der Ausfall eines einzelnen Dienstes keine anderen Dienste im System oder die Verfügbarkeit des Knotens beeinträchtigen. Außerdem kann jeder Dienst innerhalb weniger Sekunden unabhängig voneinander neu gestartet werden. Diese Designprinzipien gelten sowohl für Front-End-Dienste wie Datenprotokolle (NFS, iSCSI, FC, NVMe-oF) als auch für Back-End-Datendienste wie Neural Cache, InfiniRaid® und InfiniSnap®.

Die Datendienste werden vom Cluster-Manager (CLM) gestartet und überwacht. Er erkennt etwaige Probleme mit den Diensten und kann diese bei Bedarf neu starten. Tritt bei einem Dienst ein Fehler auf, wird er neu gestartet und führt einen Selbsttest durch, bevor er dem Cluster wieder beitritt.
Ein nicht korrekt startender Dienst tritt auch nicht dem Cluster bei, um einen Ausfall im Cluster zu vermeiden (byzantinischer Fehler). Ermittelt der Cluster-Manager einen Dienst, der mehrmals erfolglos versucht hat, auf einem bestimmten Knoten neu zu starten, stoppt er den Neustart und benachrichtigt den Infinidat-Support. Jeder Dienstausfall, ob automatisch wiederhergestellt oder nicht, wird an die Datenanalyseplattform von Infinidat zurückgemeldet, um Softwareprobleme zu erkennen und die Codequalität kontinuierlich zu verbessern.

FESTPLATTENANORDNUNG

Die Anordnung der InfiniBox-Festplatten wird durch eine patentierte Softwareinnovation namens InfiniRaid verwaltet. Es handelt sich um ein softwaredefiniertes, redundantes Array unabhängiger Festplatten (RAID), das die gesamte Datenplatzierung, den Datenschutz und die Wiederherstellung nach Ausfallszenarien steuert. InfiniRaid ist ein „entclustertes RAID", bei dem die Datenanordnung von der physikalischen Schicht getrennt ist und Tausende virtueller RAID-Gruppen verwendet werden, wodurch die Daten auf alle Laufwerke verteilt und Problemstellen vermieden werden. InfiniRaid erstellt die RAID-Gruppen so, dass sich jeweils zwei Laufwerke im System nur bis zu 2,5 % ihrer RAID-Gruppen teilen. Dieser niedrige Prozentsatz an überlappenden RAID-Gruppen bietet mehrere Vorteile:

- **Selbstheilung**: Alle potenziellen Problemstellen werden automatisch durch die Optimierung der Datenanordnung behoben.
- **Virtuelle Reservekomponenten**: Die Platzkapazität wird gleichmäßig auf alle Festplatten im System verteilt. Es gibt keine physischen Reservekomponenten zum Austausch im Betrieb, so dass beim Neuaufbau die Daten optimal umverteilt und unnötige Kosten minimiert werden. Das System verfügt über genügend Reservekapazität für bis zu zwölf Laufwerke, die bei einem F6000 ausfallen können.
- **Leistungsschutz**: Der Ausfall eines einzelnen Laufwerks (die Daten sind immer noch geschützt) führt nur zu einem RAID-Neuaufbau mit niedriger Priorität („Rebuild-1"), bei dem die Anwendungsleistung im Vordergrund steht.
- **Schnelle Wiederherstellung**: Fällt ein zweites Laufwerk aus, priorisiert das System den Neuaufbau für die 2,5 % der RAID-Gruppen, die von beiden ausgefallenen Laufwerken gemeinsam genutzt werden („Rebuild-2"), bevor es auf den Rebuild-1 mit niedrigerer Priorität zurückgreift, da es keine ungeschützten RAID-Gruppen mehr gibt.
- **InfiniSpares**: Über die garantierte Kapazität für das Äquivalent von zwölf Reservekomponenten hinaus kann InfiniBox bei Bedarf auch freie Kapazität als Reservekapazität nutzen. Diese Innovation ermöglicht den Ausfall von bis zu 100 Festplatten, ohne dass der Schutz verloren geht.

Datensicherheitsdienste

InfiniBox bietet zahlreiche Datensicherheitsdienste, mit denen Kunden ihre Vermögenswerte schützen können:

- **Snapshots**: InfiniSnap ist die InfiniBox-Funktion für Snapshots. Dieser nicht sperrende, beim Schreiben umleitende Mechanismus liefert eine gleichbleibende Performance – mit oder ohne Snapshots. Von jedem Datensatz können bis zu 1.000 Snapshots erstellt werden, die entweder schreibgeschützt (um Daten zu schützen) oder beschreibbar (für Test- und Entwicklungsumgebungen) sein können. InfiniSnap erstellt Snapshots im DRAM, sodass keine Schreibvorgänge in der Persistenzschicht erforderlich sind.
- **Unveränderliche Snapshots**: InfiniSnap ermöglicht auch unveränderliche Snapshots. Gemäß den bei der Snapshoterstellung festgelegten Anforderungen und Zeitvorgaben können sie nach dem Schreibvorgang nicht mehr bearbeitet oder gelöscht werden. Mithilfe dieser Funktion werden Daten optimal vor Ransomware-Angriffen und anderen Bedrohungen geschützt.
- **Asynchrone Replikation mit niedrigem RPO**: Mit einem Replikationsintervall von 4 Sekunden erreicht das asynchrone Replikationsmodul kontinuierlich den branchenweit niedrigsten Recovery Point Objective (RPO). Dank Nutzung von IP-Infrastruktur werden Kosten und Komplexität reduziert.
- **Synchrone Replikation**: Das synchrone Replikationsmodul bietet synchronen Schutz von Daten mit einem RPO-Wert von null und einer Speicherlatenz von unter 400 µs. Bei Problemen mit dem WAN (hohe Latenzzeit, Verbindungsunterbrechungen) fällt das synchrone InfiniBox-Replikationsmodul automatisch in den asynchronen Modus zurück. Direkt nach der WAN-Wiederherstellung werden sämtliche fehlenden Daten automatisch repliziert, und die synchrone Replikation wird ohne I/O-Unterbrechungen fortgesetzt.
- **Active/Active-Replikation**: Active/Active-Replikation mit InfiniBox-Systemen ermöglicht gleichzeitiges Lesen und Schreiben in Konsistenzgruppen über ein MAN. Sie bewahren ein externes Abbild der Speichervolumen, die wie Multipfade zum selben Volumen erscheinen. Mithilfe synchroner Replikation wird dabei zu jedem Zeitpunkt die Konsistenz der Volumen sichergestellt. Ohne Master/Slave-Beziehung sind keine zusätzlichen Roundtrips erforderlich, um Schreib-Updates für ein Volumen

auszuführen. Ein externer Lightweight-„witness“ kann auf einem eigenständigen Knoten oder auch einer virtuellen Maschine in einer Cloud gespeichert werden.

- **Gleichzeitige Replikation an einen dritten Speicherort**: Jede Konsistenzgruppe in einer Active/Active-Replikationsbeziehung kann ohne zusätzliche Leistungseinbußen zugleich auch asynchron an einen dritten Speicherort repliziert werden. Da jedes InfiniBox-System das Speichern eines zweiten Remote-Replikats jeder Konsistenzgruppe ermöglicht, kann jede auf das gleiche InfiniBox-Drittsystem – oder sogar auf ein viertes System an einem separaten Standort – repliziert werden.
- **InfiniSafe**: InfiniSafe erweitert die Fähigkeiten der InfiniBox-Produktfamilie im Bereich Cyber-Resilienz. Diese Technologie nutzt unveränderliche Snapshots, schafft einen lokale logische Trennung und Isolation, richtet eine abgeschirmte forensische Umgebung ein und ermöglicht eine nahezu sofortige Wiederherstellung nach Cyberangriffen.

Datenreduktion

InfiniBox setzt mehrere Methoden der Datenreduzierung ein, um die Speicherkosten weiter zu senken, darunter:

- **Standardmäßiges Thin Provisioning**: Alle Volumen werden standardmäßig als nicht speicherbelastend (thin) bereitgestellt. Da InfiniBox auch intelligente Kapazitätspools anbietet, kann das Risiko einer Überzuteilung/overprovisioning leicht durch das Einrichten von Warnschwellen und Notfallpuffern für den Pool gemindert werden, was wiederum die Verfügbarkeit der Anwendung schützt.
- **Nullenbasierte Rückgewinnung (Zero-Reclamation)**: Sobald die Hosts (physisch oder virtuell) Speicherplatz auf einer Festplatte (LUN) freigeben, schreiben sie Nullen in diesen Speicherplatz, entweder durch mehrfaches Kopieren desselben Datenblocks (effizienter) oder einfach durch Schreiben einzelner Nullen in diesen Speicherplatz. InfiniBox erkennt beide Fälle und entfernt diesen Speicherplatz, als ob er nie beschrieben worden wäre, was das Thin Provisioning weiter verbessert.
- **Komprimierung**: InfiniBox komprimiert Daten erst, wenn sie vom Schreib-Cache (DRAM) auf die Festplatte übertragen werden. Dadurch werden Schreibvorgänge beschleunigt (keine zusätzliche Latenzzeit aufgrund der Datenreduzierung) und gleichzeitig wird die Komprimierung flüchtiger Daten, die nach einigen Sekunden überschrieben werden, vermieden (Einsparung von CPU-Ressourcen). Die InfiniBox-Komprimierung nutzt LZ4 mit einer Blockgröße von 64 KiB, wodurch eine höhere Komprimierungsrate als bei der herkömmlichen Komprimierung in kleinen Blöcken (die in allen Flash-Arrays verwendet wird) erzielt wird.
- **Snapshots**: InfiniBox-Snapshots sind als platzsparend konzipiert. Die Kunden vermeiden so die Kapazitäts- und Leistungseinbußen einer vollständigen Kopie.

Netzwerkarchitektur

Bei allen netzwerkbasierten Diensten ist die Zugänglichkeit des Netzwerks entscheidend für die Verfügbarkeit. Insbesondere bei IP-basierten Diensten (iSCSI, NFS, SMB, asynchrone Replikation, synchrone Replikation) erwarten IT-Administratoren in der Regel, dass das Speichersystem Failover handhabt und Konfigurationsprobleme schnell überwindet. InfiniBox ist in diesem Bereich innovativ, weil es bei einem Konnektivitätsproblem ein sofortiges IP-Failover einsetzt und IP-Adressen auf Netzwerkschnittstellen verschiebt, die die entsprechenden Dienste bereitstellen können.

Sofortiges IP-Failover gilt für alle Ausfallszenarien, einschließlich Hardware (Knotenausfall, Ausfall von Ethernet-Ports/Netzwerkkarten) oder Software (Dienstausfall auf einem bestimmten Knoten). Um die Auswirkungen auf andere Dienste zu minimieren, verschiebt InfiniBox die minimale Anzahl von IP-Adressen, so dass keine IPs eines anderen Dienstes auf diesem Knoten oder IPs auf anderen Knoten verschoben werden.

InfiniBox nutzt auch virtuelle MAC-Adressen (VMAC) und ordnet jede IP-Adresse einer VMAC zu. Wenn IP-Adressen verschoben werden, verschieben sich auch die VMAC-Adressen. Dadurch entfällt die Failover-Zeit und die Konfigurationsänderung kann auf dem Switch erfolgen, ohne dass die Änderung auf jeden Host übertragen wird. Dadurch werden auch die unnötigen ARP-Probleme vermieden und die Verfügbarkeit wird erhöht.

InfiniBox verwendet intelligente Netzwerküberwachung (mit ICMP-Ping über IPv6), um potenzielle Fehlkonfigurationen zu erkennen, z. B. die versehentliche Blockierung einer Speichernetzwerkschnittstelle für den Zugriff auf ein VLAN, das für Datendienste verwendet wird. Jedes in der InfiniBox konfigurierte Netzwerk wird ständig überwacht, so dass den Speicheradministratoren oft die Frage: „Warum hat diese Anwendung den Zugriff auf den Speicher verloren?“ beantwortet wird, noch bevor sie sie selbst stellen.

Hardwarearchitektur

InfiniBox ist ein softwaredefiniertes Speichersystem, das Standard-Hardware nutzt. Als Teil des Konzepts hat Infinidat in Software investiert, um die Standard-Hardware zuverlässiger, kosteneffektiver, einfacher zu verwalten und zu unterstützen. Das wichtigste Konstruktionsprinzip ist N+2: Alle Komponenten sind mindestens dreifach redundant, um eine Zuverlässigkeit von sieben Neunen und in Verbindung mit der InfiniBox-Software eine Systemverfügbarkeit von 100 % zu erreichen.

Das InfiniBox-System wird, wie hier gezeigt, vormontiert in einem Rack geliefert:

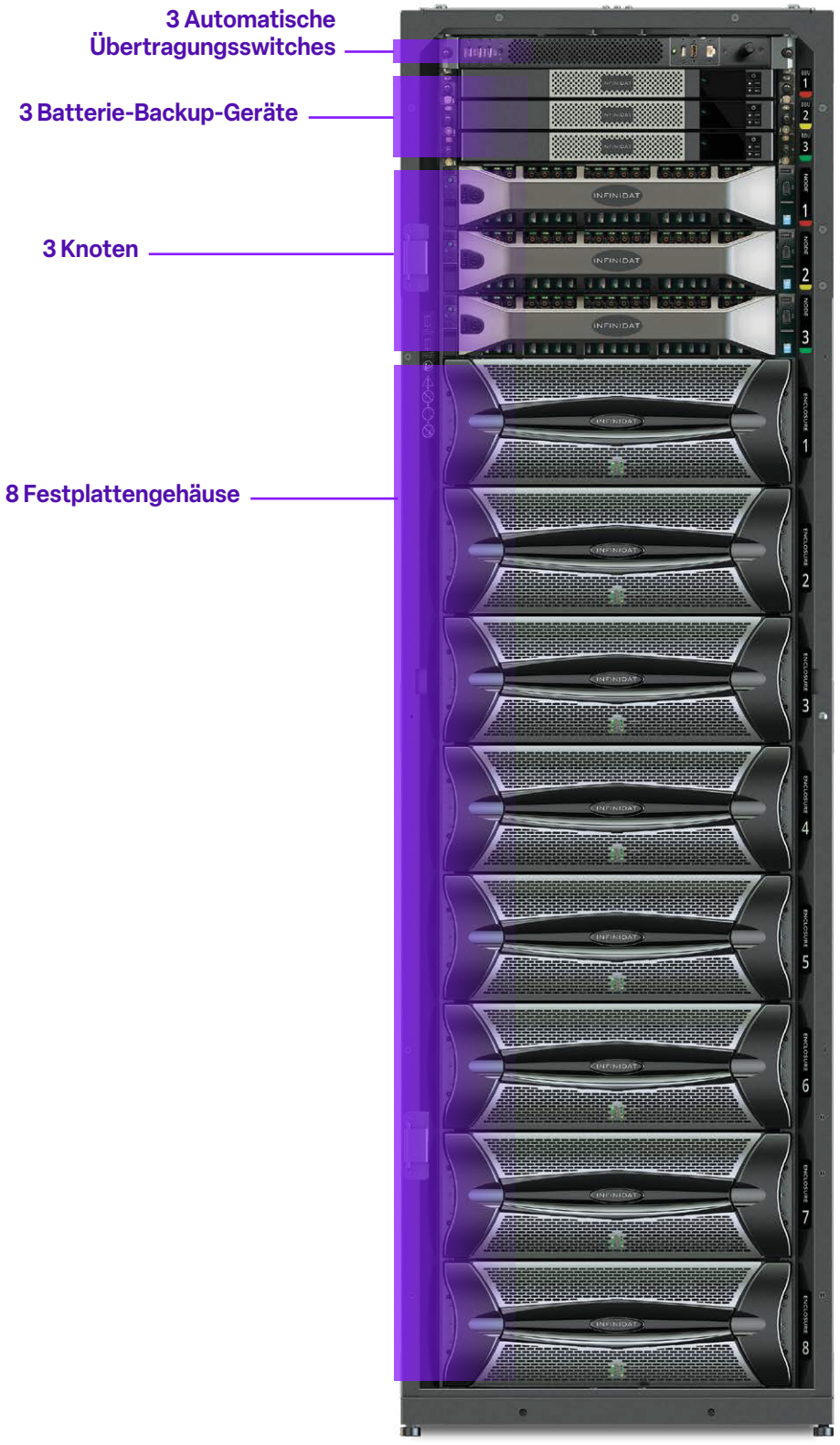


BILD 1 InfiniBox-Rack: Vorderansicht

KNOTEN

Die Knoten sind die Speicher-Controller in der InfiniBox. Die drei vollständig redundanten Knoten arbeiten in einem Active/Active/Active-Cluster, so dass I/Os-Vorgänge nahtlos über alle drei Knoten fließen können. Die Knoten sind direkt mit schnellem InfiniBand zum direkten Zugriff auf den Speicher über RDMA verbunden, was eine schnelle Replikation neuer Schreibvorgänge zwischen den Knoten mit der geringstmöglichen Latenz ermöglicht.

Bei Ausfall eines Knotens übernehmen die beiden verbleibenden Knoten dessen Aufgaben, synchronisieren alle nicht mehr replizierten Teile des Schreibcaches, um den vollständigen Datenschutz wiederherzustellen und den Betrieb unterbrechungsfrei aufrechtzuerhalten. Die N+2-Knoten-Architektur vereinfacht auch Wartungsarbeiten an einem bestimmten Knoten (z. B. den Austausch einer Komponente), da das System immer noch über zwei aktive Knoten verfügt, die die Daten schützen.

Physische Konnektivität

Front-End-Konnektivität von den Knoten zum Fabric des Kunden:

- **Fiber Channel (FC)**: Acht Ports pro Knoten, 24 Ports insgesamt. Da alle Ports aktiv sind, sieht jeder Host mehrere Pfade (mindestens einen pro Knoten, empfohlen werden zwei pro Knoten). Dank Multi-Pathing wirkt sich ein Port- oder HBA-Ausfall nur auf den einzelnen Pfad aus, nicht aber auf die Anwendungen.
- **Ethernet (Eth)-Ports**: Bis zu sechs Ports pro Knoten, 18 Ports insgesamt, die entweder Kupfer- oder Glasfaserverbindungen bieten und die Protokolle iSCSI, NVMe/TCP, NFS, SMB, synchrone Replikation und asynchrone Replikation unterstützen. Diese Ports unterstützen intelligentes IP-Failover, damit kein physischer Ausfall die Systemzugänglichkeit beeinträchtigt.

Intern sorgen die Knoten auch für die redundante Backend-Konnektivität:

- **InfiniBand (IB)-Ports**: Verwendet für die Cluster-Verbindung. Jeder InfiniBand-Ausfall, der zu einer Unterbrechung der Verbindung zwischen zwei Knoten führt, bewirkt, dass diese beiden Knoten über den dritten Knoten kommunizieren. Wird ein Knoten von den beiden verbleibenden Knoten getrennt, wird er ordnungsgemäß aus dem Cluster entfernt, bis die Unterbrechung behoben ist.
- **SAS-Ports**: Verbindung der Knoten mit allen Festplattengehäusen. Bei einem SAS-Ausfall, durch den ein bestimmter Knoten auf einigen Festplatten nicht mehr zugänglich ist, erfolgt mit dem InfiniBand der Fernzugriff auf diese Festplatten über einen anderen Knoten.

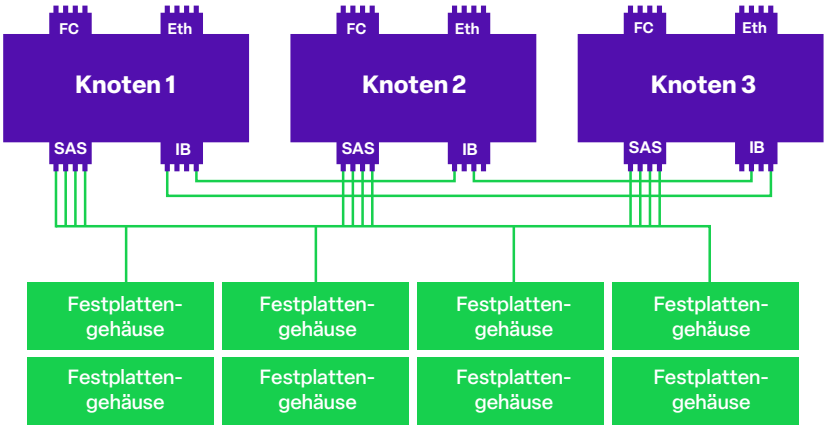


BILD 2 InfiniBox Front-End und Back-End Konnektivität

Die Knoten verfügen über redundante Stromversorgungen und werden von verschiedenen Batterie-Backup-Geräten (BBUs) gespeist, die wiederum über mehrere Stromanschlüsse versorgt werden, um auch bei Problemen mit der Stromversorgung einen unterbrechungsfreien Betrieb zu ermöglichen.

Automatische Übertragungsswitches

Die automatischen Übertragungsswitches (ATS) steuern die Stromeinspeisung in die BBUs und stellen sicher, dass die Batterie auch bei Ausfall einer der Stromquellen immer mit Strom versorgt wird. Die ATS können sofort zwischen zwei Stromquellen umschalten, wenn eine von ihnen ausfällt, und so die Stromversorgung der BBU unterbrechungsfrei aufrechterhalten.

Batterie-Backup-Geräte

Die BBUs sorgen dafür, dass die InfiniBox-Knoten auch bei kurzen Stromausfällen mit Strom versorgt werden (z. B. bis die Generatoren wieder voll aktiv sind), so dass das System nicht heruntergefahren werden muss. Sie liefern auch den Strom, um bei einem längeren Stromausfall die Daten ordnungsgemäß aus dem DRAM-Cache auf internen Platten zu sichern, so dass die InfiniBox stets ordnungsgemäße Abschaltverfahren durchführen kann. Die BBUs werden überwacht und einmal pro Woche automatisch getestet, um Batterieprobleme zu erkennen und sicherzustellen, dass das System im Falle eines echten Stromausfalls geschützt ist.

Fazit

Die einzigartige InfiniBox-Architektur überwindet die traditionellen Kompromisse zwischen Zuverlässigkeit, Leistung, Kapazität und Gesamtbetriebskosten. Die Konzentration auf Software als treibende Kraft für Speicherinnovationen macht es jetzt möglich, Lösungen zu implementieren, die sich mit der Zeit verbessern. Zum ersten Mal können IT-Organisationen im Rahmen ihres begrenzten Budgets arbeiten und gleichzeitig ihrem Unternehmen die Möglichkeit geben, neue Initiativen durchzuführen. Mit Infinidat können Unternehmen ihre wichtigsten Unternehmensdaten einfacher und kostengünstiger erfassen, speichern, analysieren und schützen, um Wettbewerbsvorteile zu erzielen.

dach@infinidat.com

WP-INFARCH-220515-DE | © 2023 Infinidat, ein Unternehmen von Lenovo

Lenovo INFINIDAT