



HPE EZMERAL DATA FABRIC

Moderne Infrastruktur für Datenspeicherung und -management



INHALT

Kurzübersicht	3
Argumente für eine neue Dateninfrastruktur	3
Auswahlkriterien für eine moderne Dateninfrastruktur	3
HPE Ezmeral Data Fabric	4
Definition einer Data Fabric	4
Grundlagen der HPE Ezmeral Data Fabric	4
Der Nutzen der HPE Ezmeral Data Fabric	4
Beispielvorteil: Optimierung von Architektur und Workflows	5
Funktionsweise der HPE Ezmeral Data Fabric	7
Offener Datenzugriff über mehrere APIs mit globalem Namespace	7
Jenseits der Dateien: integrierte Tabellen und Ereignis-Streams	9
Skalierbarkeit, Zuverlässigkeit und Leistung	11
Datenmanagement auf Plattformebene	12
Datenbewegung auf Plattformebene	13
Globale Datenbereitstellung	17
Cloud-unabhängige Datenportierbarkeit	18
Integriertes Daten-Tiering	19
Integrierte Sicherheit	20
HPE Ezmeral Data Fabric Dateisystem	21
Nutzung der HPE Ezmeral Data Fabric	21
Ressourcen	22



KURZÜBERSICHT

Die Komplexität, die geografischen Dimensionen und der Umfang moderner Datensysteme stellen den Betrieb vor enorme Herausforderungen, die die Entwicklung ausbremsen, eine stabile Produktion einschränken und eine Hürde für Skalierbarkeit und Innovation darstellen können. Es ist jedoch möglich, diesen Herausforderungen erfolgreich zu begegnen. Mit einem effektiven Design auf Grundlage der richtigen unterstützenden Technologien lassen sich viele dieser Herausforderungen vermeiden. Deshalb wird eine neue Infrastrukturlösung für Datenspeicherung und -management benötigt.

Eine moderne Dateninfrastruktur kann als Grundlage zur Vereinheitlichung dienen, mit der Sie Systeme zusammenstellen können, die weniger komplex sind, bessere Leistung erbringen und hochgradig skalierbar sind. Gleichzeitig sind sie ausfallsicherer als Systeme, die aus verschiedenen Einzellösungen zusammengestellt werden. Die HPE Ezmeral Data Fabric ist eine einzigartige Datentechnologie, die mit genau diesem Ziel entwickelt wurde.

Als vollständig softwarebasierte und hardwareunabhängige Lösung liefert die Data Fabric eine globale Dateninfrastruktur vom Edge bis zur Cloud, um Daten bedarfsgemäß zu erfassen, zu speichern, zu verwalten und zu bewegen und die Datenanforderungen umfangreicher Analyse- und KI-Anwendungen zu erfüllen. Sie kann von containerisierten oder nicht containerisierten Anwendungen genutzt werden. Die Data Fabric bietet flexible Datenzugriffsmethoden, sodass sowohl ältere als auch moderne Anwendungen direkt auf die in der Data Fabric gespeicherten Daten zugreifen können.

Dieser Bericht erläutert die Funktionsweise der HPE Ezmeral Data Fabric als vereinheitlichende Dateninfrastruktur zur Speicherung und Verwaltung von Daten in umfangreichen Systemen. Die Data Fabric hilft bei der unternehmensweiten Industrialisierung datenintensiver Projekte. Sie optimiert Leistung und Sicherheit auf kosteneffiziente Weise ohne zusätzliche Belastung der IT-Teams.

Folgende Konzepte werden behandelt:

- Warum eine moderne Dateninfrastruktur wichtig ist
- Was ist die HPE Ezmeral Data Fabric?
- Wie die HPE Ezmeral Data Fabric Architekturen und Workflows optimiert
- Funktionsweise der HPE Ezmeral Data Fabric

ARGUMENTE FÜR EINE NEUE DATENINFRASTRUKTUR

Das schnelle Wachstum von Datensätzen und die zunehmende Vielzahl von Techniken, um aus Daten Mehrwert zu gewinnen, veranlassen Unternehmen häufig dazu, komplexe Ansammlungen aus verschiedenen Systemen zusammenzustellen, um die Anforderungen von geschäftskritischen, analytischen sowie KI- oder Machine-Learning-Anwendungsfällen zu erfüllen. Dieser aus Einzellösungen bestehende Ansatz erhöht Kosten und Komplexität und erfordert häufig die Nutzung spezialisierter IT-Ressourcen, die schwer anzupassen oder einem neuen Zweck zuzuführen sind, wenn sich die Unternehmensanforderungen ändern. Herkömmliche Softwarelösungen eignen sich vielleicht gut für die Aufgaben, für die sie ursprünglich vorgesehen waren, lassen sich jedoch häufig nicht über ihre ursprüngliche Auslegung hinaus skalieren, ohne ihre Leistung oder Zuverlässigkeit zu beeinträchtigen. Darüber hinaus kann die Skalierung traditioneller Lösungen oft äußerst kostspielig werden.

Es wird ein neuer Ansatz für Datenspeicherung und -management benötigt. Die Lösung ist eine moderne Dateninfrastruktur, die dafür entwickelt wurde, Datenspeicherung und -management einheitlich in Ihrem gesamten Unternehmen zu optimieren – mit dem großen Umfang von heute und der nötigen Skalierbarkeit für zukünftige Anforderungen.

AUSWAHLKRITERIEN FÜR EINE MODERNE DATENINFRASTRUKTUR

Woran erkennt man die erforderlichen Leistungsmerkmale, damit eine Technologie als vereinheitlichende Dateninfrastruktur dienen kann? Hier finden Sie eine Zusammenfassung der **wichtigsten Auswahlkriterien**:

- Nachweislich zuverlässiger und hochverfügbarer Betrieb in der Produktion
- Auf Exabyte-Niveau skalierbar
- Hervorragende Leistung im großen Maßstab und bei Multi-Tenant-Nutzung
- Ausfallsicherheit mit Selbstreparatur, eigenständigem Ausgleich und ohne Single Point of Failure
- Komfortables und effektives Management von Daten, Anwendungen und Teams mit Zugriffskontrollausdrücken auf mehreren Ebenen
- Erreichbarkeit für eine Vielzahl von Anwendungen und Tools (containerisiert und nicht containerisiert; Analysen und KI; älter und modern)
- Erreichbarkeit über Cloud-native Anwendungen mit standardmäßiger CSI
- Unterstützung für geografische Verteilung für Edge-, Core-, Cloud- und Hybrid-Cloud-Bereitstellungen mit Datenbewegung auf Plattformebene und Remote-Datenzugriff über globalen Namespace



- Effizientes Daten-Tiering und Löschungskodierung zur Optimierung der Ressourcennutzung
- Integrierte Sicherheit mit Unterstützung für die Verschlüsselung von Daten während der Übertragung und auf dem Laufwerk

Die HPE Ezmeral Data Fabric ist eine moderne Dateninfrastruktur, die diese Auswahlkriterien erfüllt. Dies wird in der nachfolgenden Erläuterung näher besprochen.

HPE EZMERAL DATA FABRIC

Um zu verstehen, wie die HPE Ezmeral Data Fabric wesentliche Vorteile für die Anforderungen moderner umfangreicher Systeme bietet, denken Sie zunächst darüber nach, was eine Data Fabric ist.

Definition einer Data Fabric

Der Begriff **Data Fabric** wird ganz unterschiedlich ausgelegt. Manche verwenden ihn in einem eingeschränkten Kontext und meinen damit nur eine Abstraktions- und Managementebene. Unsere Definition einer **Data Fabric** ist hingegen viel umfangreicher – insbesondere unter Berücksichtigung von Datenspeicherung, -management und -bewegung. HPE Ezmeral Data Fabric ist ein Beispiel für diese umfassendere Bedeutung von „Data Fabric“.

Grundlagen der HPE Ezmeral Data Fabric

Die [HPE Ezmeral Data Fabric](#) ist eine hochgradig skalierbare Datentechnologie, die von Grund auf als vereinheitlichende Dateninfrastruktur für das gesamte Unternehmen entwickelt wurde. Sie bietet eine wirtschaftliche und leistungsfähige Datenerfassung und -speicherung, bequemes Datenmanagement und effiziente Datenbewegung vom Edge bis zum Core, von lokalen Rechenzentren bis zu Cloud-, Multi-Cloud- und Hybrid-Cloud-Bereitstellungen. Dank einzigartiger Technologie ermöglicht sie eine Kombination und Vereinfachung von Systemen und somit ein neues Design für die Systemarchitektur. Die HPE Ezmeral Data Fabric unterstützt den Datenzugriff sowohl für Legacy-Anwendungen als auch für moderne Analyse- oder KI-Anwendungen – alles auf dem gleichen System, mit der gleichen Sicherheit und unter der gleichen Administration.

Die Kerntechnologie der Data Fabric ist ein verteiltes Datenspeicherungs- und -managementsystem mit einer überzeugenden Erfolgsgeschichte in Produktionssystemen in einer Vielzahl von umfangreichen Installationen. **Mit der HPE Ezmeral Data Fabric erhalten Sie Dateien, Tabellen und Ereignis-Streams zusammen in einer umfassenden Dateninfrastruktur.**

Die Data Fabric ist aus einer Technologie entstanden, die früher als MapR Data Platform bekannt war. (MapR Technologies wurde im Jahr 2019 von HPE übernommen.) Die [HPE Ezmeral Data Fabric](#) ist eine hardwareunabhängige reine Softwarelösung, die als Bestandteil des neuen [HPE Ezmeral Softwareportfolios](#) verfügbar ist. Darüber hinaus ist sie als Kern-Datenebene für die [HPE Ezmeral Container Platform](#) verfügbar.

Der Nutzen der HPE Ezmeral Data Fabric

Die HPE Ezmeral Data Fabric bietet eine Vielzahl von Vorteilen im Vergleich zur Zusammenstellung umfangreicher Systeme aus Einzellösungen.

Einige der wichtigsten Vorteile:

- Nahtlose Skalierung von Daten und Anwendungen ohne Skalierung von IT-Mitarbeitern
- Zuverlässige und einfachere Erfüllung von Service-Level-Garantien
- Konsolidierung von Systemen zur Vermeidung von Wildwuchs und Senkung der Kosten
- Ermöglichung von Daten vom Edge bis zur Cloud und in Hybrid Cloud-Bereitstellungen
- Aufhebung der Notwendigkeit einer Neugestaltung von Datensystemen für neue Projekte oder Ansätze
- Schutz dank integrierter globaler Sicherheit und fein abgestimmter Administrationsrichtlinien

Sehen wir uns an, wie diese Vorteile zusammenarbeiten, um Nutzen zu schaffen. Wie in [Abbildung 1](#) illustriert, können verschiedene Workloads, einschließlich Legacy-Anwendungen und moderner KI- und Analyseanwendungen, Daten nutzen, die auf dem gleichen System gespeichert und verwaltet werden. Das Potenzial einer geografisch verteilten Bereitstellung trägt auch zu der Fähigkeit der Data Fabric bei, Workflows zu vereinfachen, Architekturen zu optimieren und somit Wildwuchs zu reduzieren, selbst wenn das System größer wird.



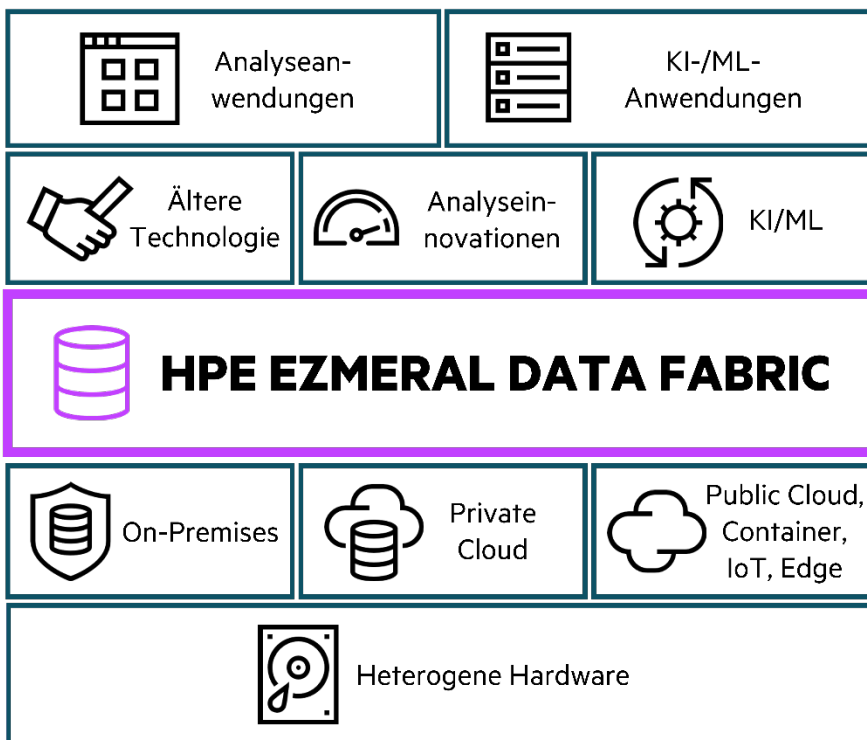


ABBILDUNG 1. HPE Ezmeral Data Fabric für Datenspeicherung und das Management der Datenlogistik

Die gemeinsame Nutzung von Daten in einem Multi-Tenant-System reduziert nicht nur die Kosten, sondern verbessert auch die Zusammenarbeit und bietet Analysten und Data Scientists ein umfassenderes Verständnis der Daten. Systemadministratoren und Benutzer können effizienter arbeiten, weil die Data Fabric Optionen für das Management von Datenlogistik und Zugriffskontrollen auf Plattformebene bietet. Das automatisierte Datenmanagement und die Selbstreparaturfunktionen der Data Fabric gewährleisten Leistung und Zuverlässigkeit, selbst in großen Umfängen. Ein IDC-Bericht untersuchte den geschäftlichen Nutzen der HPE Ezmeral Data Fabric und befragte dazu Unternehmen, die die Data Fabric für den geschäftskritischen Betrieb bereitgestellt hatten. Der Bericht kam zu dem Ergebnis, dass diese Unternehmen 85 % weniger ungeplante Ausfallzeiten sowie eine 33%ige Erhöhung der Produktivität ihrer Datenanalyseteams erzielten.

Kurz gesagt: Die HPE Ezmeral Data Fabric ist branchenweit einzigartig, weil sie eine leistungsstarke einheitliche Datengrundlage auf einem System bildet, die die Produktivität der IT und die Datensicherheit erhöht und gleichzeitig Komplexität und Kosten senkt.

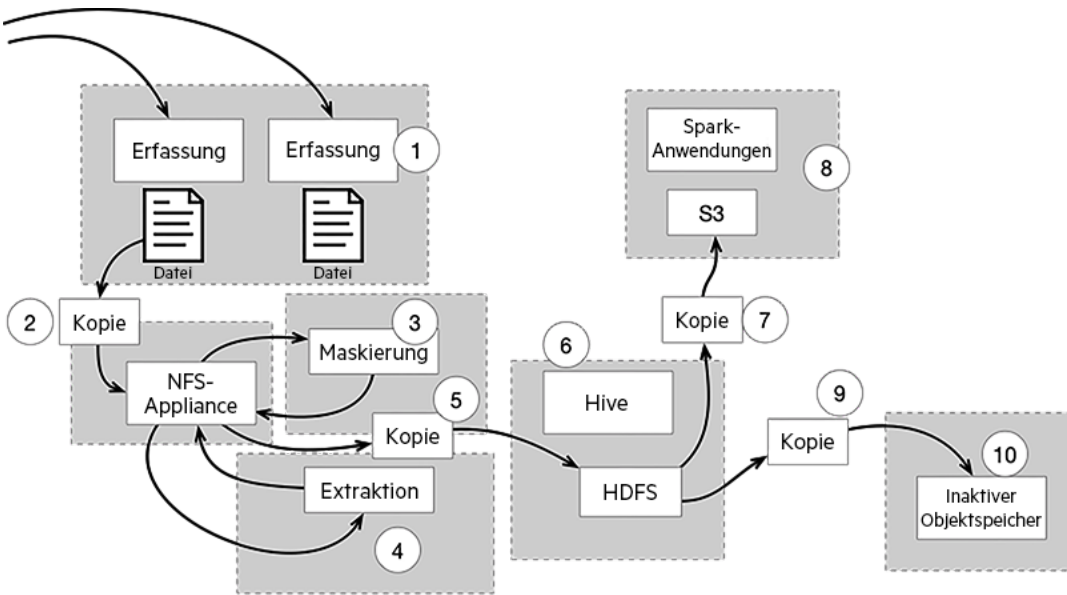
BEISPIELVORTEIL: OPTIMIERUNG VON ARCHITEKTUR UND WORKFLOWS

Um den Nutzen der HPE Ezmeral Data Fabric besser zu verstehen, ist es nützlich, ein Beispiel für ihre überzeugenden Vorteile zu betrachten. Denken Sie darüber nach, wie sich ein typischer Anwendungsfall mit der Einführung der Data Fabric verändern würde.

Einer der spürbarsten Vorteile bei der Verwendung der HPE Ezmeral Data Fabric ist die Vereinfachung, die durch die Reduzierung der Verbreitung mehrerer Einzellösungen für Datenspeicherung und -management erzielt wird. Beachten Sie, dass die Data Fabric nicht nur für spezialisierte Big-Data-Anwendungen vorgesehen ist. Vielmehr erfüllt die HPE Ezmeral Data Fabric die Datenanforderungen von großen Projekten im ganzen Unternehmen.

Diese Vereinfachung wird unter anderem dadurch ermöglicht, dass die Data Fabric über verschiedene APIs offen aufgerufen werden kann. Somit kann eine Vielzahl von Anwendungen und Tools direkt auf die Daten in der Data Fabric zugreifen, ohne Anpassungen oder spezielle Bibliotheken zu erfordern. Abbildung 2 (Teil A und B) illustriert, wie die HPE Ezmeral Data Fabric eine typische Architektur vereinfachen kann. In Abbildung 2A basiert jeder grau schattierte Block auf einem anderen System oder Tool. Der Workflow wird durch das System aus Pfeilen verdeutlicht. Beachten Sie, wie viele Schritte in diesem Workflow das Kopieren großer Datensätze zwischen Systemen erfordert – ein zeitraubender und potenziell unnötiger Prozess.

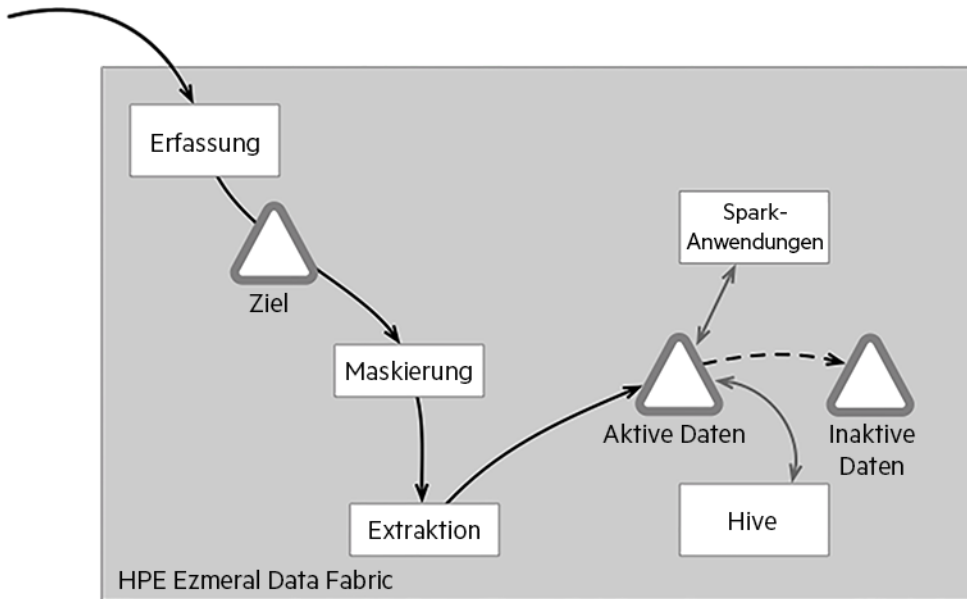




Teil A: Ursprüngliche Architektur

ABBILDUNG 2A. System ohne Data Fabric: typischer komplexer Workflow und Architektur für ein System aus Einzellösungen für Daten

Vergleichen Sie Workflow und Architektur des Systems ohne Data Fabric aus Abbildung 2A mit dem deutlich optimierten System auf Basis der HPE Ezmeral Data Fabric in Abbildung 2B.



Teil B: Vereinfachte Architektur

ABBILDUNG 2B. Optimierter Workflow und Architektur auf HPE Ezmeral Data Fabric

Mit einer Data Fabric als unternehmensweiter einheitlicher Datenspeicherungs- und -managementinfrastruktur werden viele Schritte im Workflow, beispielsweise das Kopieren der Daten, unnötig. Der Workflow kann jetzt auf einem einzigen System durchgeführt werden (Data Fabric, in Abbildung 2B durch den einzelnen großen schattierten Block illustriert).



Der Workflow (Schritt 1 bis 10) in Abbildung 2A ist eine vereinfachte Darstellung eines typischen realen Anwendungsfalls, in dem verschiedene Systeme, z. B. lokale Dateien, NFS-Farm, HDFS usw., zum Speichern von Daten verwendet werden. Jedes dieser Datensysteme mag sinnvoll gewesen sein, als es zum System hinzugefügt wurde, doch aus moderner Sicht sorgen sie nur für ein unnötiges Maß an Komplexität. Das Management so vieler unterschiedlicher Systeme ist eine unangemessene Last, die einen Großteil der Arbeitszeit eines IT-Teams einnimmt. Beispielsweise entsteht fast die Hälfte der gesamten Komplexität des Systems ohne Data Fabric durch das Kopieren von Daten von System zu System. Zusätzlich führen unterschiedliche Sicherheitsmodelle auf diesen Systemen zu Sicherheitslücken, die von einem Angreifer ausgenutzt werden könnten.

Vergleichen Sie diese Komplexität mit der Einfachheit von Abbildung 2B, die illustriert, wie der gleiche Workflow mit der HPE Ezmeral Data Fabric aussehen würde. Schritte zur Datentransformation stehen noch aus, doch es ist nicht mehr notwendig, Daten von System zu System zu kopieren. Rechenressourcen können flexibel und bedarfsgemäß von verschiedenen Prozessen genutzt werden, weil jeder Prozess auf jeder Maschine über die gleichen Pfadnamen und über eine beliebige API auf alle in der HPE Ezmeral Data Fabric gespeicherten Daten zugreifen kann. Dieser universelle Datenzugriff beseitigt viele Schritte und separate Systeme, die in der ursprünglichen Architektur (Abbildung 2A) erforderlich waren. Ein umfassendes, kryptografisch basiertes Sicherheitssystem macht den universellen Zugriff sicher.

Darüber hinaus kann die HPE Ezmeral Data Fabric dank ihrer integrierten Funktionen den gesamten Lebenszyklus der Daten verwalten – von unformatierten Daten bis hin zur Maskierung und geordneten Verarbeitung von Daten, die für verschiedene Anwendungen bereit sind. Auf Wunsch handhabt die Data Fabric sogar die Platzierung von Daten in Offlinespeicher. All das geschieht ohne Änderung der Wahrnehmung der Daten durch Anwendungen. Ebenso wichtig ist, dass die IT Zeit und Aufwand sparen kann, indem sie globale Daten über ein einziges System verwaltet, das herauf- oder herunterskaliert werden kann, um Anwendungen zu unterstützen, ohne den Management-Workload zu verändern.

FUNKTIONSWEISE DER HPE EZMERAL DATA FABRIC

Nachdem wir die Grundprinzipien der Data Fabric und ihre Vorteile erläutert haben, befassen wir uns nun mit bestimmten entscheidenden Funktionen, die benötigt werden, damit eine Dateninfrastruktur auch bei sehr großen Maßstäben eine grundlegende, einheitliche Datenebene bilden kann.

Beachten Sie, dass die HPE Ezmeral Data Fabric ein hochgradig skalierbares verteiltes Datenspeichersystem ist, das als echtes, vollständig lese- und schreibfähiges Dateisystem entwickelt wurde. Überraschenderweise enthält die Data Fabric neben Dateien auch Tabellen und Ereignis-Streams als native Objekte erster Klasse im verteilten Dateisystem. (Ereignis-Streams werden auch als Nachrichten-Streams bezeichnet).

Dies ist ein einzigartiges Design. Anstelle von separaten Komponenten, die über Konnektoren funktionieren, wurden die **Dateien, Tabellen und Streams der Data Fabric zusammen als eine Technologie entwickelt, die als Bestandteil eines Systems ausgeführt wird**. Dieses einheitliche Datensystem vereinfacht Architektur, Workflows und Datenmanagement und steigert die Leistung.

Offener Datenzugriff über mehrere APIs mit globalem Namespace

Im Hinblick auf den Datenzugriff unterscheidet die HPE Ezmeral Data Fabric sich durch folgende Faktoren von anderen Lösungen:

- Universeller Datenzugriff über mehrere offene APIs, die eine Vielzahl von Anwendungen unterstützen, die alle direkt auf die Daten in der Data Fabric zugreifen können
- Globaler Namespace, der Rechenzentren, Edge und Cloud umfasst

Hier erfahren Sie, wie diese Kombination von Funktionen die Zusammenstellung einfacherer und leistungsstarker umfangreicher Systeme ermöglicht.

Datenzugriff über mehrere APIs

Durch die Bereitstellung eines offenen Datenzugriffs über mehrere APIs ermöglicht die HPE Ezmeral Data Fabric Legacy-Anwendungen die direkte Zusammenarbeit mit modernen Analyse- und KI-Anwendungen – bei sicherer gemeinsamer Nutzung der gleichen Daten auf dem gleichen System.

Wer mit anderen Big-Data-Technologien vertraut ist, wird überrascht sein, dass die HPE Ezmeral Data Fabric die Verwendung des gesamten Spektrums an bekannten Linux®-Befehlen sowie die unveränderte Ausführung von Apache Spark- oder Apache Hadoop-Programmen erlaubt. Auch Kubernetes-basierte Anwendungen können direkt auf die Daten auf der Data Fabric zugreifen und mit ihnen arbeiten, weil diese einen integrierten Container Storage Interface (CSI-)Treiber bietet.

Die HPE Ezmeral Data Fabric unterstützt eine Reihe von offenen Datenzugriffsprotokollen, einschließlich POSIX-basierter APIs, Big Data APIs wie HDFS sowie Cloud-basierter APIs wie Amazon Simple Storage Service (AWS S3). Dank dieses flexiblen Datenzugriffs können Entwickler und Analysten das Tool oder die Sprache ihrer Wahl nutzen (d. h. Java, Python, Apache Spark, Apache Hive, SQL). Data Scientists können KI- und Machine-Learning-Tools für den direkten Zugriff auf Daten auf der Data Fabric nutzen, ohne die Daten zuerst auf ein spezielles System kopieren zu müssen. Der universelle Datenzugriff durch verschiedene Anwendungen wird in Abbildung 3 illustriert. Die in Abbildung 3 illustrierten Anwendungen greifen trotz unterschiedlicher APIs über den gleichen Pfadnamen auf Daten zu.



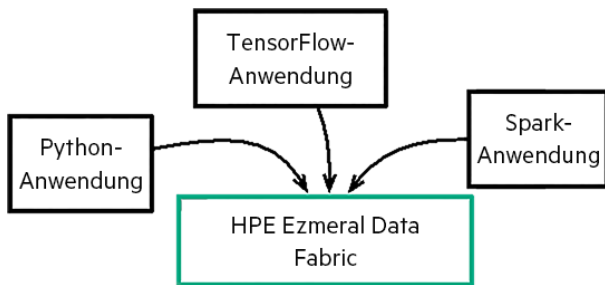


ABBILDUNG 3. Direkter Zugriff auf Daten auf der HPE Ezmeral Data Fabric durch Anwendungen, die unterschiedliche APIs nutzen und mit verschiedenen Tools entwickelt wurden

Selbst vertraute Linux-Befehle können genutzt werden, um die Daten auf der Data Fabric zu durchsuchen. Der Screenshot in Abbildung 4 illustriert, was passiert, wenn der Linux-Befehl `ls` zur Anzeige eines Data-Fabric-Verzeichnisses verwendet wird.

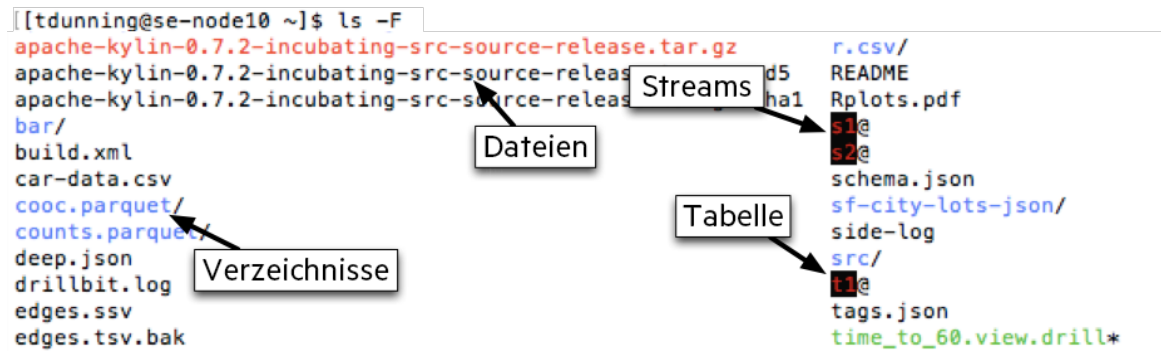


ABBILDUNG 4. Linux-Befehl `ls` zur Anzeige der Inhalte eines Verzeichnisses der HPE Ezmeral Data Fabric

Die Illustration in Abbildung 4 bedeutet, dass die Data Fabric den Datenzugriff sowohl auf bekannte traditionelle Weise als auch über moderne APIs ermöglicht. Beachten Sie außerdem, dass das Verzeichnis Unterverzeichnisse, Dateien, Tabellen und Streams enthält. Die HPE Ezmeral Data Fabric dient als echte vereinheitlichende Datenebene. Der Datenzugriff über mehrere APIs ist eine der vielen Möglichkeiten, auf die die Data Fabric die Vorteile der Mandantenfähigkeit in die Tat umsetzt.

Ein weiterer Vorteil der Data Fabric als vereinheitlichende Dateninfrastruktur ist die Bereitstellung eines globalen Namespace – ein Merkmal mit beträchtlichen Vorteilen.

Globaler Namespace

Mit der HPE Ezmeral Data Fabric befinden sich alle Dateien, Tabellen und Ereignis-Streams der Data Fabric in einem einzigen globalen Namespace, selbst wenn sie sich auf separaten Clustern an unterschiedlichen Standorten befinden. Ein globaler Namespace bewirkt, dass eine Anwendung auf Daten auf einem lokalen Data-Fabric-Cluster oder remote auf Daten auf einem Data-Fabric-Cluster an einem anderen Standort zugreifen kann, einschließlich Edge, On-Premises, Rechenzentren oder in Cloud- und Multi-Cloud-Bereitstellungen (Abbildung 5).



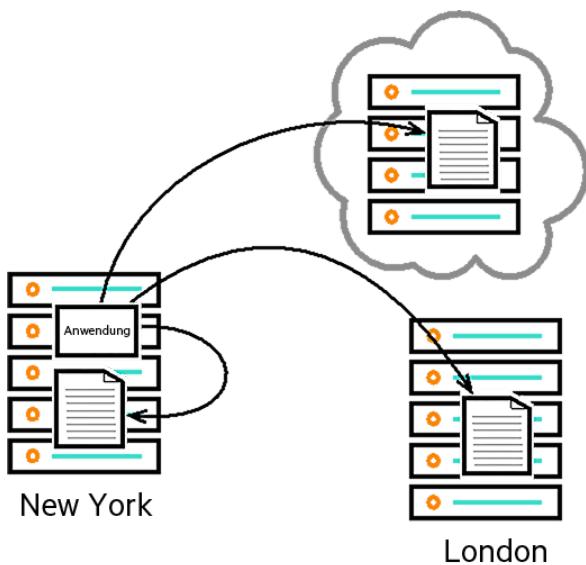


ABBILDUNG 5. Durch den globalen Namespace der HPE Ezmeral Data Fabric ermöglichter lokaler und Remote-Datenzugriff

Stellen Sie sich die Wirkung dieses globalen Namespace vor. Entwickler, Analysten und Data Scientists können Daten auf Remote-Clustern referenzieren, als wären sie lokal gespeichert. Für sie wirkt es, als befänden sich die Daten gleich vor ihnen. Der einfache Zugriff vereinfacht Anwendungen und erleichtert mehreren Anwendungen die gemeinsame Arbeit an den gleichen Daten. Er sorgt auch für eine deutliche Vereinfachung des Designkonzepts großer Systeme.

Zusätzlich hilft der globale Namespace bei der Aufgabentrennung, sodass diejenigen, die Programme und Modelle erstellen, sich darauf konzentrieren können, wie ihre Anwendungen die Daten nutzen, statt sich um die Datenlogistik zu sorgen. Die Aufgabentrennung entlastet wiederum Systemadministratoren, indem sie sich stärker auf die Administration der Plattform und weniger auf Anliegen auf Anwendungsebene konzentrieren können.

Gemeinsam sind der universelle Datenzugriff über mehrere offene APIs und der globale Namespace vom Edge bis zur Cloud der Schlüssel zur Vereinheitlichung der Dateninfrastruktur durch die HPE Ezmeral Data Fabric. Zu den Vorteilen zählen eine flexible Tool- und Sprachenauswahl, die Möglichkeit, Legacy-Anwendungen und moderne Anwendungen auf der gleichen zugrundeliegenden Datenebene auszuführen, sowie ein effizienteres Datenmanagement. Diese Merkmale machen echte Mandantenfähigkeit zu einer geschickten Designoption.

Jenseits der Dateien: integrierte Tabellen und Ereignis-Streams

Ein überraschender Aspekt der HPE Ezmeral Data Fabric besteht darin, dass NoSQL-Tabellen und Ereignis-Streams (Nachrichten-Streams) als Objekte erster Klasse in das Dateisystem integriert sind. Diese kombinierte Funktion für skalierbare, verteilte Dateien, Tabellen und Ereignis-Streams – alle auf der gleichen technologischen Basis entwickelt – ist einzigartig.

Die Tabellen- und Ereignis-Stream-Funktionen der Data Fabric sind nicht nur für sich genommen äußerst leistungsfähige Technologien – integrierte Tabellen und Ereignis-Streams machen die Data Fabric zudem effektiver als eine vereinheitlichende Dateninfrastruktur. Benutzern (Entwicklern, Analysten, Data Scientists) und IT-Teams bieten die Tabellen und Ereignis-Streams der Data Fabric ein ungewöhnliches Maß an Flexibilität und Komfort. Hier erfahren Sie, wie.

Dateien, Tabellen und Streams nutzen die gleiche Sicherheit und das gleiche Administrationssystem, selbst an geografisch verteilten Standorten. Sie sind im gleichen globalen Namespace enthalten und Richtlinien können zusätzlich zum Management von Zugriffsberechtigungen auf Volume-Ebene auf Stream- oder Tabellenebene angewendet werden (weitere Informationen dazu finden Sie im Abschnitt „Datenmanagement auf Plattformebene“). Darüber hinaus können Tabellen und Ereignis-Streams direkt und nahezu in Echtzeit zwischen Clustern repliziert werden, wie im Abschnitt „Andere Formen der Datenbewegung“ dieses Dokuments erläutert wird. Sie können auch den Zustand von containerisierten, von Kubernetes orchestrierten Anwendungen in jeder Form, z. B. Dateien, Ereignis-Streams oder Tabellen, in der Data Fabric beibehalten.

Integrierte Ereignis-Streams der Data Fabric

Daten in Ereignis-Streams der Data Fabric können über die Apache Kafka-API abgerufen werden (obwohl die Streams der Data Fabric ganz anders implementiert sind) und erhalten Funktionen, die über Kafka hinausgehen. Erfahren Sie als Erstes, wie Streams als Nachrichtentechnologie funktionieren.



Daten aus vielen Quellen (Produzenten) werden in [Themen](#) geschrieben und viele – potenziell Tausende oder noch mehr – Themen werden zu einem einzigen Data-Fabric-Stream kombiniert. Verbraucheranwendungen lesen Nachrichten aus Streams (statt der für Verbraucher übertragenen Ereignisdaten), wobei eine Nachricht durch das Lesen nicht gelöscht wird. Somit können Verbraucher auch lange nach dem Schreiben der Ereignisse in Themen hinzugefügt werden. So erhält der Entwickler ein beträchtliches Maß an Flexibilität, weil die Produzenten und Verbraucher von Nachrichten hochgradig voneinander entkoppelt und somit unabhängig entwickelt und bereitgestellt werden können. Um einen Stream zu erstellen, muss lediglich eine Datei erstellt werden, und um neue Themen zu einem Data-Fabric-Ereignis-Stream hinzuzufügen, können sie einfach beim Schreiben einer Nachricht verwendet werden. Das macht Streams bequem und effizient für die Verwendung durch Entwickler.

Nützlich ist ebenfalls, dass jede Partition in einem Thema über den gesamten Data-Fabric-Cluster verteilt und nicht auf eine einzige Maschine eingeschränkt wird. Durch diese Eigenschaft werden Data-Fabric-Ereignis-Streams hochgradig skalierbar. Tatsächlich ist es sinnvoll, die Lebenszeit für Daten in einem Stream bei Bedarf auf Monate oder sogar Jahre festzulegen. Dieser Ansatz kann für einen reproduzierbaren, ereignisbezogenen Verlauf verwendet werden – eine nützliche Funktion für Audit- oder Compliance-Anforderungen.

Jeder Ereignis-Stream – potenziell mit Tausenden oder sogar Millionen von Themen – entspricht im Wesentlichen einem ganzen Kafka-Cluster. Und ein einziger Data-Fabric-Cluster kann bei Bedarf Millionen von Streams unterstützen. Viele Streams können selbst zwischen Clustern zu einem einzigen Stream repliziert werden. Diese Merkmale machen die Data Fabric insbesondere in industriellen IoT-Anwendungsfällen nützlich, in denen Daten aus vielen Edge-Quellen gesammelt und entweder teilweise verarbeitet oder als unformatierte Daten in zentrale Rechenzentren bewegt werden müssen. Dies wird im Abschnitt [„Datenbewegung auf Plattformebene“](#) im Detail betrachtet.

Integrierte Tabellen der Data Fabric

[Zwei Arten von Tabellen sind in die Data Fabric integriert.](#) Diese sind:

- Binäre Tabelle
- Dokumententabelle

Beide Arten von Tabellen sind genauso einfach zu erstellen wie eine Datei. Wie in der Diskussion zum globalen Namespace erwähnt, können [Data-Fabric-Tabellen](#) in jedem beliebigen Verzeichnis gespeichert werden, in dem Sie eine Datei speichern würden. Data-Fabric-Tabellen verfügen ebenso wie Dateien über Pfadnamen und die Genehmigungen nutzen die gleiche Art von Zugriffskontrollausdrücken. Das bedeutet, dass Volumes genutzt werden können, um Dateien, Tabellen und Verzeichnisse gemeinsam auf konsistente Weise zu verwalten.

Diese Eigenschaften der integrierten Tabellen in der Data-Fabric-Infrastruktur ermöglichen eine unkomplizierte Verwendung von Tabellen in Anwendungen. Die Integration von Tabellen (und Ereignis-Streams) in das verteilte Dateisystem der Data Fabric ermöglicht Ihnen die Erstellung einer breiten Palette an Anwendungen auf einem einzigen Cluster, ohne separate Cluster für Tabellen oder Ereignis-Streams installieren, verwalten oder skalieren zu müssen. Natürlich können die Dateien in einer Data Fabric auch als zugrundeliegender Speicher für andere, herkömmlichere Arten von Tabellen verwendet werden, beispielsweise relationale oder NoSQL-Datenbanken.

Darüber hinaus bietet die Art und Weise, auf die Tabellen in die Data Fabric implementiert sind, nicht nur extreme Skalierbarkeit mit hervorragender Leistung, sondern auch ein hohes Maß an Flexibilität. Neue Tabellen oder Indizes können hinzugefügt und die Daten in Tabellen können ganz einfach neu strukturiert werden, selbst im großen Maßstab. Moderne Unternehmen müssen schnell handeln und auf veränderte Situationen oder Geschäftsziele reagieren können, was die durch Data-Fabric-Tabellen ermöglichte Flexibilität zu einem wichtigen Resultat der Verwendung der Data Fabric macht.

Ein zusätzlicher Vorteil von integrierten Tabellen besteht darin, dass sie durch das gleiche Management auf Plattformebene verwaltet werden wie Dateien. Fein abgestimmte Zugriffskontrollen mithilfe von ACEs können auf Tabellenebene angewandt oder von dem Volume übernommen werden, das die Tabelle enthält. ACEs können auch unterhalb der Tabellenebene angewandt werden: Spalten in binären Data-Fabric-Tabellen und Felder in Data-Fabric-Dokumententabellen (beachten Sie dabei, dass ACEs anders und deutlich leistungsfähiger sind als Zugriffskontrolllisten [ACLs], die in Apache HBase verwendet werden).

Binäre Data-Fabric-Tabellen

Binäre Tabellen in der HPE Ezmeral Data Fabric werden über die Apache HBase-API abgerufen, die Schlüssel und Werte von nicht typisierten Byte-Arrays zulässt. Binäre Data-Fabric-Tabellen sind ebenso skalierbar, zuverlässig, verfügbar und leistungsfähig wie die zugrundeliegende Fabric und werden seit Jahren ohne Ausfallzeiten in geschäftskritischen Systemen eingesetzt. Data-Fabric-Tabellen unterliegen nicht den Verdichtungsverzögerungen, die beim Ausführen von Apache HBase selbst zu bewältigen sind. Beispielsweise verwendet ein großer Regierungskunde in Asien seit Jahren binäre Data-Fabric-Tabellen für Authentifizierungsprozesse in einem nationalen Ausweissystem – ohne Ausfallzeiten, selbst bei Hardware-Wartungs- und -Austauschvorgängen. Dieses System weist extreme SLAs für Zuverlässigkeit und Verfügbarkeit auf, die mit der vorherigen Implementierung des Systems mit HBase unmöglich einzuhalten waren.

Data-Fabric-Dokumententabellen

Die integrierten Dokumententabellen der Data Fabric sind in vielen modernen Situationen die bevorzugte Option für neue Anwendungen, weil sie verschachtelte Datenstrukturen im JSON-Stil unterstützen. Diese dokumentenorientierten Tabellen werden über die Open-Source-API [OJA](#) abgerufen.



Dokumentenorientierte Tabellen bieten ein hohes Maß an Flexibilität, wodurch es besonders einfach wird, Änderungen vorzunehmen oder die Tabelle mit neuen Datenkategorien zu aktualisieren. Beispielsweise könnte ein Online-Händler mithilfe dieser Tabellen seinen Produktkatalog speichern. Was passiert, wenn er eine neue Art von Produkt mit neuen Arten von Informationen hinzufügen möchte? Bei traditionellen Datenbanken kann eine solche Änderung zeitraubend sein und sogar die Hinzufügung neuer Tabellen oder Ausfallzeiten erfordern, um die Datenbank neu aufzubauen. Im Gegensatz dazu können Datensätze mit dokumentenorientierten Tabellen denormalisiert werden, um verschachtelte Strukturen einzubeziehen, und strukturelle Änderungen können schnell auf eine Weise vorgenommen werden, wie sich eine REST API im Laufe der Zeit weiterentwickeln kann.

Data-Fabric-Dokumententabellen unterstützen außerdem [sekundäre Indizes](#) mit der gleichen Skalierbarkeit wie die Tabellen selbst. Die Verwendung von Indizes ist für Anwendungen transparent.

Native Erfassung von Datenänderungen (Change Data Capture, CDC)

Eine nützliche Funktion der Data Fabric ist die Erfassung aller Änderungen an den Inhalten einer Data-Fabric-Dokumententabelle. Hierzu können Sie jede Tabelle so konfigurieren, dass Inhaltsänderungen an einen damit verknüpften [Erfassungs-Stream für Datenänderungen](#) übermittelt werden. Dank dieser CDC-Streams können Änderungen überwacht, untersucht oder zur Aktualisierung von zusammenfassenden Tabellen genutzt werden.

Skalierbarkeit, Zuverlässigkeit und Leistung

Nur allzu häufig kommt es bei der Skalierung von Systemen zu unangenehmen Kompromissen zwischen Skalierbarkeit und Zuverlässigkeit oder Leistung. Möglicherweise haben die Systeme in einem bestimmten Umfang gut funktioniert, werden aber aufgrund der steigenden Anforderungen großer Datensätze schon bei einem geringfügig größeren Umfang unzuverlässig. Die Verantwortlichen für das Design und den Aufbau des Systems beginnen häufig mit dem Einsatz eines Flickenteppichs von Problemumgehungen zur Aufrechterhaltung der Zuverlässigkeit, nur um dann eine Verschlechterung der Leistung (oder umgekehrt) festzustellen. Typischerweise wird aus dem System ein Mosaik, bei dem jeder Teil auf separaten Datenplattformen liegt. Doch diese Kompromisse lassen sich vermeiden. Die HPE Ezmeral Data Fabric ermöglicht den Ausbau von Systemen auf große Umfänge auf der gleichen Dateninfrastruktur, während SLAs für Zuverlässigkeit, Verfügbarkeit, Latenz und Gesamtleistung weiterhin erfüllt werden.

Ein entscheidendes Designmerkmal für die Erfüllung dieser Kombination von Anforderungen ist die vollständige Verteilung der Metadaten für Verzeichnisse, Dateien, Tabellen und Streams im gesamten Cluster mit der HPE Ezmeral Data Fabric. Verteilte Metadaten sind leistungsstark, aber selten, weil sie schwierig zu implementieren sind. Sie bieten Schutz vor Datenverlusten oder mangelnder Verfügbarkeit, indem Single Points of Failure und Hotspots vermieden werden, die andere verteilte Dateisysteme einschränken. Zusätzlich sorgt die Art, auf die verteilte Daten in die HPE Ezmeral Data Fabric implementiert werden, für Ausfallsicherheit, weil die Metadaten bei ihrer Verteilung repliziert werden. Darüber hinaus verbessert dieses Design die Leistung, indem es eine hohe Parallelität in Daten- und Metadatenpfaden ermöglicht.

Zusätzliche Zuverlässigkeit entsteht durch [automatische Datenreplikation](#). Die Replikationsniveaus können konfiguriert werden. Standardmäßig werden drei Replikate erstellt, die über möglichst viele Fehlerdomänen verteilt werden. Wird ein Laufwerk so beschädigt, dass ein Replikat verloren geht, erstellt die Data Fabric automatisch eine neue Kopie aus den zwei verbleibenden Replikaten. Dabei läuft das System ungestört weiter, selbst wenn die Replikation im Hintergrund durchgeführt wird. Anwendungen greifen transparent auf die verbleibenden Datenreplikate zu. Für Anwendungen (oder Entwickler oder Analysten) gibt es in der Regel keine spürbaren Auswirkungen. Diese [Selbstreparaturfunktion der HPE Ezmeral Data Fabric](#) hilft bei der Gewährleistung der Zuverlässigkeit und Leistung selbst in sehr großen Clustern.

Auch durch die hochentwickelten **Lastenausgleichsmechanismen** der Data Fabric werden Leistung und Zuverlässigkeit gestärkt. Diese Funktion ist besonders wichtig, wenn mehrere Anwendungen versuchen, auf die gleichen Daten zuzugreifen. Um Staus zu verhindern, ermöglicht die HPE Ezmeral Data Fabric das Lesen von mehreren Replikaten. Soweit verfügbar, nutzt die Data Fabric außerdem auch mehrere Netzwerkverbindungen. Zusätzlich werden Staus vermieden, indem Daten automatisch bewegt werden, um Hotspots oder Staus zu vermeiden und den Datenverkehr zwischen Netzwerkpfaden ins Gleichgewicht zu bringen.

Der Umfang ist nicht nur von der Datenmenge abhängig

Beachten Sie bei der Suche nach einer Dateninfrastruktur, die Leistung und Zuverlässigkeit im großen Maßstab gewährleistet, dass es beim [Umfang nicht nur auf die Menge an Bytes ankommt](#), die in einem System gespeichert und verwaltet werden. Eine weitere Herausforderung beim Umfang ist die Bewältigung von zahlreichen Objekten. Die Arbeit mit Hunderten von Millionen, Milliarden oder sogar noch mehr Dateien kann die Dateninfrastruktur überfordern, wenn sie nicht für einen solchen Umfang ausgelegt ist. Die HPE Ezmeral Data Fabric ist für die Bewältigung solcher Umfänge vorgesehen. Kunden mit einer Data Fabric arbeiten in der Regel mit Milliarden, teilweise sogar Billionen von Dateien.

Umfang oder Skalierbarkeit

Letztendlich kommt es nicht nur auf den derzeitigen Umfang an, sondern auch darauf, wie gut die [Dateninfrastruktur Erweiterungen bewältigen kann](#). Benutzer der HPE Ezmeral Data Fabric erweitern ihren Datenbestand häufig um 20 bis 50 % pro Jahr – jedes Jahr –, weshalb es hilfreich ist, wenn die Dateninfrastruktur dieses Wachstum auch bewältigen kann. Einige Systeme sind innerhalb von fünf Jahren oder noch weniger um mehrere Größenordnungen gewachsen. Ein kleiner HPE Ezmeral Data Fabric Cluster von 50 Terabyte kann auf 500 Petabyte anwachsen, indem einfach Hardware hinzugefügt wird. Die Selbstreparaturfunktionen der Fabric, die der Fabric die Bewältigung von Hardwareausfällen ermöglichen, begünstigen auch derartiges Wachstum, ohne den Betrieb zu unterbrechen und ohne das System neu zusammenstellen zu müssen.

Eine wichtige Anforderung zur Unterstützung der Skalierbarkeit ist Datenmanagement auf Plattformebene, das im nächsten Abschnitt dieses Dokuments erläutert wird.



Datenmanagement auf Plattformebene

Damit Sie von den Vorteilen einer vereinheitlichenden Datenebene im gesamten Unternehmen profitieren können, ist es entscheidend, dass die Dateninfrastruktur ein effizientes Datenmanagement auf geeigneten Ebenen für ein großes und komplexes System bietet. Datenmanagement auf Plattformebene ist nötig, um die IT zu entlasten – sowohl in großem als auch bei zunehmendem Umfang. Datenmanagement-Mechanismen müssen Systemadministratoren genügend Möglichkeiten bieten, um Systeme beliebiger Größe zu verwalten und die Kontrolle auf angemessene Weise an die Eigentümer der Anwendung zu übertragen. Kann das gut umgesetzt werden, wird es machbar und effizient, mit Millionen oder Billionen von Dateien und einer großen Anzahl an Anwendungen auf einem einzigen System zu arbeiten.

Data-Fabric-Volumes für effizientes Datenmanagement

Ein wichtiger Faktor für die Gewährleistung einer effizienten Datenorchestrierung auf Plattformebene durch die HPE Ezmeral Data Fabric ist ein Management-Konstrukt, das als Daten-Volume bezeichnet wird. Volumes in der Data Fabric sind die Datenmanagement-Superkraft der Data Fabric.

Ein Daten-Volume in der HPE Ezmeral Data Fabric verhält sich ähnlich wie ein Verzeichnis mit speziellen Datenmanagement-Funktionen, indem die Größe automatisch an die Menge der erfassten Daten angeglichen wird. Es ähnelt eher einem Beutel als einer starren Kiste, weil die Daten-Volumes keinen Platz einnehmen, bis sie gefüllt werden. Anschließend wachsen sie, wenn ihnen Daten hinzugefügt werden. Dieses „beutelartige“ Verhalten macht die Vorhersage von Daten-Volumes unnötig, sodass kein Speicherplatz verschwendet wird.

Abbildung 6 illustriert das Konzept des Datenmanagements auf Plattformebene durch die Kontrolle der Platzierung von Data-Fabric-Volumes. Es werden mehrere Volumes gezeigt und die Replikation der Daten in jedem Cluster wird farblich gekennzeichnet. Ein Cluster kann aus vielen Volumes bestehen, von etwa einem Dutzend bei einem kleinen Cluster bis hin zu Millionen bei großen Data-Fabric-Clustern, und auf jedem Volume können sich Millionen von Daten (und Replikaten) befinden.

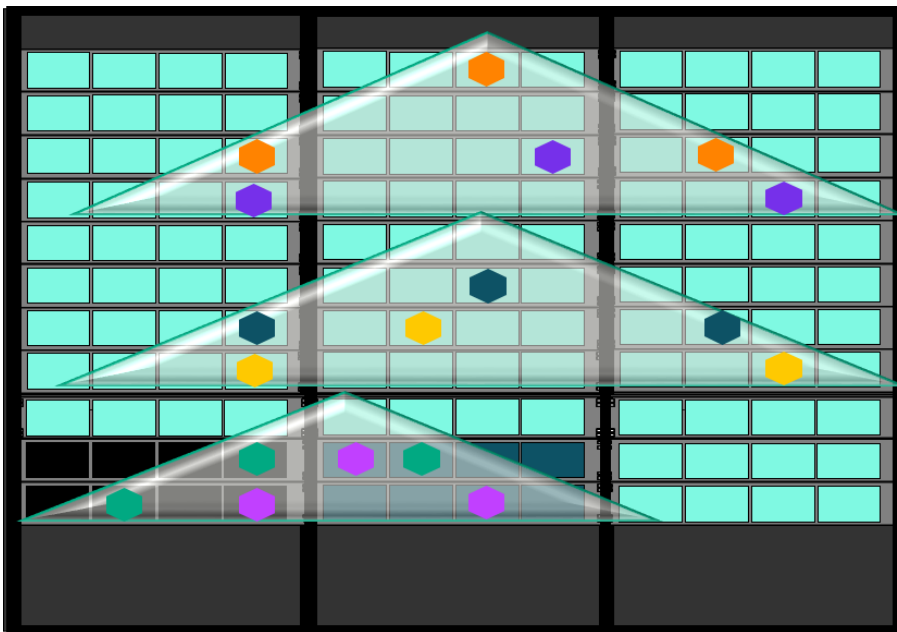


ABBILDUNG 6. Datenmanagement auf Plattformebene mit Data-Fabric-Volumes

In Abbildung 6 umfasst ein Data-Fabric-Volume (dargestellt durch ein durchsichtiges Dreieck) standardmäßig den Cluster und platziert Datenreplikate (farbige Sechsecke) für mehr Zuverlässigkeit auf verschiedenen Maschinen.

Die meisten Datenmanagement-Richtlinien werden auf Volume-Ebene festgelegt und beinhalten insbesondere die konkrete Datenplatzierung. In Abbildung 6 wird es zwei Volumes erlaubt, sich über alle Server zu erstrecken, die Daten enthalten, während ein drittes Volume eingeschränkt ist, damit die Daten auf oder in der Nähe von spezialisierter Hardware (dunklere blaue Rechtecke links unten in Abbildung 6) platziert werden.

Daten werden in einer Data Fabric platziert, indem eine Platzierung ermittelt wird, die mehrere Bedingungen erfüllt und gleichzeitig die Balance maximiert. Lokalitätsbedingungen können definiert werden, um Datenreplikate in bekannten Fehlerdomänen auszugleichen. Zusätzlich können auf einigen Speichergeräten Kennzeichnungen festgelegt werden und Volumes können auf Geräte gezwungen werden, deren Leistung der jeweiligen Kennzeichnung entspricht. Algorithmen zur Datenplatzierung müssen **beide** Anforderungen erfüllen. Beispielsweise könnten Daten in unterschiedlichen Fehlerdomänen, z. B. verschiedenen Racks, platziert werden. NVMe-Speichergeräte in der Nähe von GPUs können mit Kennzeichnungen versehen werden, sodass die Platzierung von Trainingsdaten in der Nähe von maschinellen Lernprogrammen erzwungen werden kann. Abbildung 7 illustriert, wie diese Datenplatzierung unterhalb der Maschinenebene funktionieren könnte.



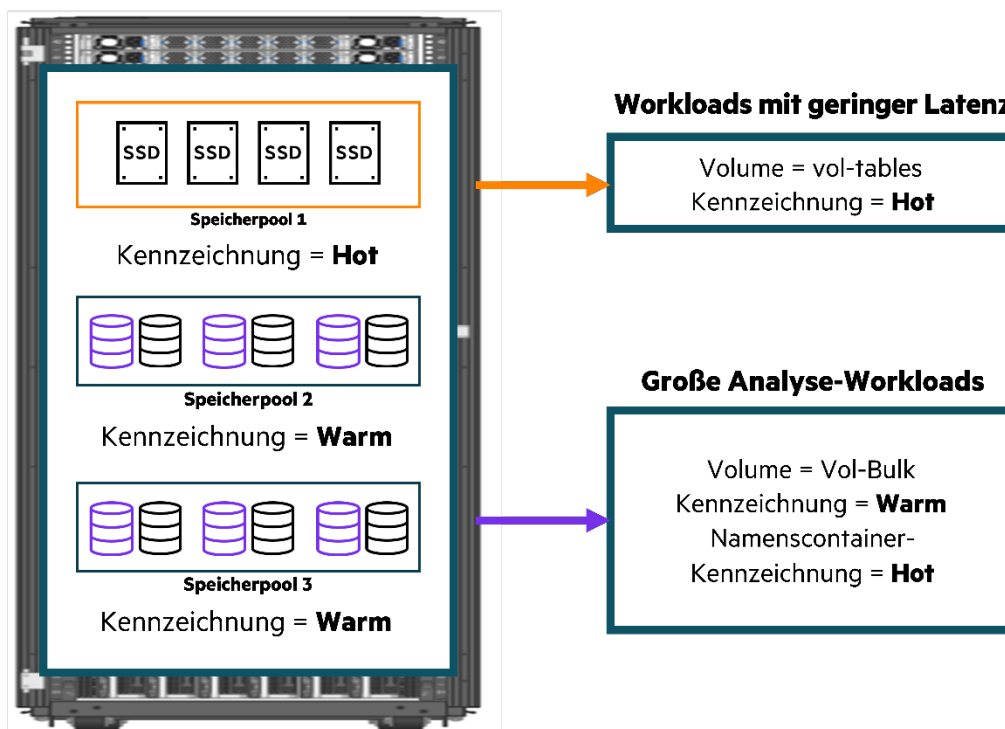


ABBILDUNG 7. Fein abgestimmte Datenplatzierung in der HPE Ezmeral Data Fabric

Durch eine Richtlinie auf Volume-Ebene wird eine Angleichung der Latenzanforderungen von Anwendungsdaten an den Medientyp ermöglicht, indem einfach die richtige Kennzeichnung angewendet wird. Die einfache Zuweisung von Richtlinien für die Datenplatzierung mithilfe von Data-Fabric-Volumes bedeutet, dass latenzempfindliche Daten auf schnelleren Speichermedien bleiben können, während große analytische Datensätze auf langsameren Medien platziert werden können, sodass Ressourcennutzung und Leistung optimiert werden. Darüber hinaus kann den Daten und Replikaten auf einem Volume eine Platzierungsrichtlinie zugewiesen werden, dem Namenscontainer (die Basis für verteilte Metadaten) eine andere.

Fein abgestimmte Kontrolle des Datenzugriffs unterstützt Mandantenfähigkeit

Quoten und Zugriffsbeschränkungen können optional auf Volume-Ebene festgelegt werden, um die Höchstmenge an Daten zu kontrollieren, die ein Volume (oder eine Volume-Gruppe gemeinsam) beinhalten kann, und zu bestimmen, wer darauf zugreifen kann. Diese Funktion gewährleistet, dass Teams und Benutzer einen sicheren Zugriff auf die Daten bekommen, allerdings **nur** auf die Daten, die sie benötigen. Diese Art des Schutzes ist in einem mandantenfähigen System entscheidend – und Mandantenfähigkeit ist eine Voraussetzung für Skalierbarkeit, ohne die IT skalieren zu müssen. Weitere Einzelheiten zur Sicherheit in der HPE Ezmeral Data Fabric werden im Abschnitt „Integrierte Sicherheit“ dieses Berichts erläutert.

Snapshots und Spiegelung

Data-Fabric-Volumes bilden auch die Grundlage für echte zeitpunktbasierte Snapshots und für Spiegelung. Snapshots und Spiegelungen können mithilfe eines automatisierten Zeitplans oder manuell erstellt und ihre Lebensdauer entsprechend festgelegt werden.

Der primäre Verwendungszweck von Snapshots besteht im Schutz der Daten vor menschlichen oder Anwendungsfehlern. Außerdem sind Snapshots eine hervorragende Möglichkeit zur Handhabung der Datenversionierung, einer besonders wertvollen Funktion für datenwissenschaftliche Teams, weil Data Scientists häufig eine exakte Version der zum Trainieren von KI- und maschinellen Lernmodellen verwendeten Daten aufbewahren müssen.

Snapshots können schnell und einfach erstellt werden und bieten Effizienz bei der Ressourcennutzung. Sie referenzieren gemeinsam die Originaldaten und verbrauchen daher nur eine zu vernachlässigende Speicherkapazität, bis diese Daten geändert werden. Die vom Snapshot geteilten Daten werden durch das Hinzufügen neuer Daten nicht beeinflusst.

Darüber hinaus unterstützen Volumes Spiegelung. Spiegelung stellt einen Mechanismus für die Bewegung von Daten zwischen Clustern und verschiedenen geophysikalischen Standorten bereit. Beispielsweise ist Spiegelung bei der Einrichtung eines sekundären Rechenzentrums im Rahmen eines Disaster-Recovery-Plans nützlich. Die Details der Spiegelung mit der HPE Ezmeral Data Fabric werden im nächsten Abschnitt erläutert.

Datenbewegung auf Plattformebene

Die HPE Ezmeral Data Fabric erleichtert die Zusammenstellung großer Systeme, indem es Benutzern und Anwendungen ermöglicht wird, die Bewegung von Daten innerhalb von und zwischen Clustern aufzurufen und zu steuern. Dies wird auf verschiedene Arten bewerkstelligt, die für Benutzer bemerkenswert einfach sind. Intern sind sie jedoch ziemlich ausgeklügelt.



Datenbewegung auf Plattformebene ist ein enormer Vorteil. Geografisch weit entfernte Cluster können eine konsistente Perspektive auf die Welt erhalten. Die Bewegung auf Plattformebene ist ein zentraler Faktor, um es der Data Fabric zu ermöglichen, Cluster in der Cloud, vor Ort oder an Edge-Standorten abzudecken. Die Integration dieser Datenbewegung in Anwendungen ist fehleranfällig und resultiert häufig in mehreren Implementierungen der gleichen Basisfunktion.

Da die Data Fabric Daten auf mehreren Maschinen speichert, ist die Bewegung der Daten zwischen diesen Maschinen ein zentraler und unvermeidbarer Bestandteil des Normalbetriebs. Neben dieser natürlichen Bewegung ermöglicht die HPE Ezmeral Data Fabric Benutzern auch die Festlegung und Steuerung der Datenbewegung auf mehrere Arten, beispielsweise in Form von Datenspiegelung und der direkten, Cluster-übergreifenden Replikation von Data-Fabric-Tabellen und Ereignis-Streams.

Datenbewegung durch Spiegelung

Der geradlinigste zu spezifizierende Mechanismus für die Datenbewegung in einer Data Fabric ist die [Volume-Spiegelung](#). Bei der Spiegelung werden Änderungen an Objekten und Metadaten in einem Quell-Volume auf ein Ziel-Volume angewendet, das sich in der Regel in einem anderen Cluster befindetet. Der Spiegelungsprozess beginnt, wenn das System einen Snapshot auf Quell- und Ziel-Volume erstellt. Alle Änderungen, die seit dem vorherigen Snapshot auf dem ursprünglichen System vorgenommen wurden, werden daraufhin auf den Ziel-Snapshot angewendet. Jetzt kann das Ziel-Volume auf atomare Weise in den neuen Zustand versetzt werden.

Die Spiegelung mit der HPE Ezmeral Data Fabric ist hocheffizient im Hinblick auf die Ressourcennutzung, weil nach der ursprünglichen Übertragung nur geänderte Daten bewegt werden. Alle weiteren Aktualisierungen geschehen schrittweise. Spiegelungen werden in der Regel in stündlichen Intervallen eingeplant und die Ziel-Snapshots werden gemäß einem Ablaufzeitplan aufbewahrt. Der Spiegelungsvorgang kann auch manuell oder über ein Skript aufgerufen werden.

Spiegelungen bieten eine effiziente Möglichkeit, große Datenmengen von einem Cluster zum anderen zu bewegen, und zwar unabhängig davon, ob sich diese Cluster im gleichen Rechenzentrum, an verschiedenen geografischen Standorten oder zwischen On-Premises-Clustern und Cloud-Bereitstellungen befinden, wie in [Abbildung 8](#) illustriert. Da Spiegelungen auf Volume-Ebene stattfinden, werden alle Dateien, Tabellen und Ereignis-Streams auf einem Volume zusammen gespiegelt. Spiegelung wird zur Verbindung von Unternehmensprozessen an verschiedenen Standorten oder zum Bewegen von Daten zwischen Edge- und Core-Clustern verwendet, wie es im Abschnitt [„Spiegelung für Edge-Datenübertragung“](#) dieses Berichts erläutert wird.

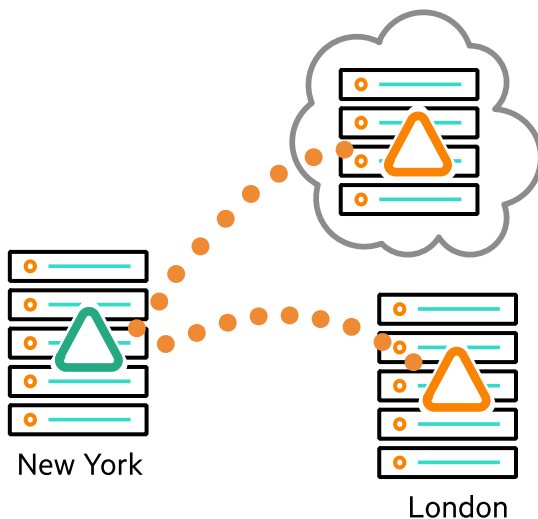


ABBILDUNG 8. Spiegelung der Data Fabric für Datenbewegung

Die Spiegelung über die Data Fabric ist eine effiziente Möglichkeit, Daten zu bewegen. Alle Daten auf dem Quell-Volume werden auf ein Spiegel-Volume auf einem Remote-Cluster kopiert, alle darauffolgenden Aktualisierungen geschehen schrittweise.

Spiegelungen werden auch häufig bei der [Vorbereitung auf Disaster Recovery](#) eingesetzt. Daten werden aus einem primären Rechenzentrum auf ein sekundäres System gespiegelt. Da die Data Fabric mehrere Vorgänge gleichzeitig unterstützen kann, wird der Wiederherstellungs-Cluster häufig als Sandbox für Entwickler oder Data Scientists genutzt, die Hardware, die andernfalls nahezu untätig wäre, sinnvoll einsetzen.

Datenbewegung ohne Störungen

Die Verschiebung von Bytes von Punkt zu Punkt wirkt zunächst unwichtig. So passt beispielsweise ein Kilobyte in ein einzelnes Netzwerkpaket. Megabytes können einfach durch einen Prozess bewegt werden – durch einfaches Schreiben in einen anderen. In beiden Fällen können einfache Netzwerkprotokolle wie TCP/IP Datenverluste korrigieren, indem sie Pakete nach Bedarf erneut übermitteln, sodass für die Bewegung dieser Datenmenge nur eine Netzwerkverbindung geöffnet und die Daten geschrieben (bzw. auf Empfängerseite gelesen) werden müssen. Doch bei riesigen Umfängen ist die Situation viel komplizierter und anspruchsvoller. Außerdem sind Datenbewegungen nicht der einzige laufende Prozess: Viele Anwendungen werden auf den Quell- und Ziel-Clustern ausgeführt und dürfen dabei nicht vom Datenbewegungsprozess gestört werden. Zusätzlich führt das System in der Regel einen kontinuierlichen Neuabgleich und eine erneute Replikation der Daten durch.



Ein Teil des Problems bei der Datenübertragung im großen Maßstab ist, dass Netzwerkverbindungen jederzeit unterbrochen werden können. Diese Unterbrechung kann sich sowohl auf die Übertragung als auch auf Bestätigungen auswirken, indem sie Verzögerungen oder eine Schließung der Verbindung verursacht. Bei einem tatsächlichen Spiegelungsprozess können buchstäblich Tausende solcher Verbindungen auf nur einer Netzwerkverbindung in Verwendung sein. Zusätzlich ist jede Verbindung nicht nur für Netzwerkunterbrechungen anfällig, sondern es können auch die Festplattenlaufwerke auf beiden Seiten ausfallen oder durch andere Vorgänge während der Spiegelung ausgelastet sein. Keine Anwendung, kein Netzwerkgerät, kein Betriebssystem und kein Festplattenlaufwerk verfügt über den nötigen Kontext, um diese Datenströme global zu handhaben, insbesondere wenn die aggregierte Übertragung von einem Cluster zum anderen berücksichtigt wird. Nur eine Data Fabric ist in der Lage, die Laufwerk-zu-Netzwerk-zu-Laufwerk-Perspektive zu bewältigen, um eine störungsfreie Datenbewegung zu gewährleisten, selbst bei enormen Umfängen. Abbildung 9 zeigt eine vereinfachte Darstellung dieser Situation in Form einer einzelnen Cluster-zu-Cluster-Verbindung.

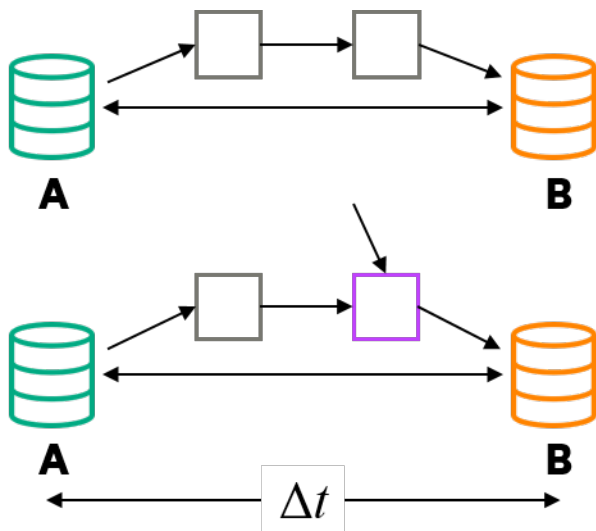


ABBILDUNG 9. Vereinfachte Darstellung der Auswirkungen von Staus bei der Datenbewegung zwischen Cluster A und Cluster B

Während der Spiegelung nutzt die HPE Ezmeral Data Fabric zeitliche Änderungen bei der vollständigen Übertragung, um Störungen (violettes Quadrat) zu erkennen und die Datenbewegung entweder zurückzuhalten oder sie auf andere Ströme umzuleiten. Im Betrieb überwacht die Data Fabric die vollständigen Übertragungszeiten von vielen Verbindungen gleichzeitig, um alle möglichen Staus oder Störungen zu erkennen. Wird eine solche Störung in einem Strom festgestellt, werden andere Ströme so gut wie möglich um die Probleme herumgeleitet, damit alle staufreien Teile des Systems möglichst produktiv bleiben.

Die Algorithmen, die Datenströme wie diesen in der HPE Ezmeral Data Fabric steuern, haben die erfreuliche Eigenschaft, dass sie nicht nur die verfügbare Kapazität der Netzwerkverbindung vollständig auslasten, sondern auch wenig oder keinen erkennbaren Druck auf gleichzeitige Ströme auf Netzwerk- oder Speichergeräteebene ausüben. Das bedeutet, dass Spiegelungen die bestehende Netzwerknutzung nicht unterbrechen und keine erkennbare Last auf dem Quell- oder Ziel-Cluster verursachen, selbst wenn ein Cluster groß und der andere klein ist oder wenn die Quelle untätig, aber das Ziel beschäftigt ist.

Neben der reinen Übertragung von Bytes von einer Festplatte zu einer entfernten Festplatte wird die Koordination der Bewegung von Millionen von Dateien zwischen Dutzenden bis Tausenden von Servern zu einem ebenso großen Problem wie die bloße Bewegung der Daten. Ein weiteres Problem ist die absolute Gewährleistung der Konsistenz und Richtigkeit der Ergebnisse. Es ist nicht unmöglich, diese Probleme zu lösen, selbst außerhalb der Data-Fabric-Implementierung, doch eine vollständige Lösung ist für Anwendungsentwickler nicht machbar. Und eine Verdoppelung dieser Bemühungen in jeder Anwendung ist verschwenderisch. Deshalb ist es sinnvoll, vom Design der HPE Ezmeral Data Fabric zu profitieren, das diese Datenübertragungsprobleme auf Plattformebene löst.

Spiegelung für Edge-Datenübertragung

Edge Computing erfordert in der Regel einen platzsparenden Cluster mit Funktionen zur Erfassung großer Datenmengen und zur lokalen Verarbeitung. Eine weitere Herausforderung beim [Edge Computing](#) ist die Bewegung von Daten, Anwendungen und Modellen zwischen Core-Clustern und dem Edge. Im Core-Cluster können weitere Analysen sowie KI/maschinelles Lernen angewendet werden, um Muster in einem globalen Umfang zu erkennen. Es muss auch die Möglichkeit geben, Daten, Modelle oder Anwendungen vom Core zurück zum Edge oder sogar von anderen Edge-Standorten zu bewegen.

Mit der HPE Ezmeral Data Fabric als einheitliche Dateninfrastruktur, die sich bis zum Edge erstreckt, werden diese Herausforderungen effizient bewältigt. Die Data Fabric stellt eine Dateninfrastruktur auf den Edge-Clustern, um Daten zu erfassen und zu speichern und Anwendungen für die lokale Verarbeitung zu unterstützen. Darüber hinaus kann die Data Fabric Daten in einer schrittweisen Spiegelung zwischen Edge-Clustern und Core bewegen.

Dank Spiegelungen können Petabytes an Daten aus Millionen von Dateien mühelos zuverlässig bewegt werden. Tatsächlich gelingt es einigen Benutzern der HPE Ezmeral Data Fabric auf diese Weise, täglich Dutzende Petabyte an Daten zu erfassen und zu analysieren. Ein reales Beispiel für die Verwendung der Data Fabric zur Erfassung und Bewegung großer Mengen an geografisch verteilten Daten finden Sie in der Fallstudie [„Schnellere Entwicklung autonomer Fahrzeuge durch einfachen Zugriff auf globale Daten“](#).



Die Datenbewegung zwischen Edge-Installationen und Core-Rechenzentren wird auch durch die Cluster-übergreifende Replikation von Tabellen und Ereignis-Streams ermöglicht, wie im nächsten Abschnitt dieses Dokuments beschrieben. Abbildung 10 zeigt eine konzeptionelle Darstellung des Layouts großer industrieller IoT-Systeme. Dank der Data Fabric kann das Gesamtsystem global lernen, aber lokal handeln.

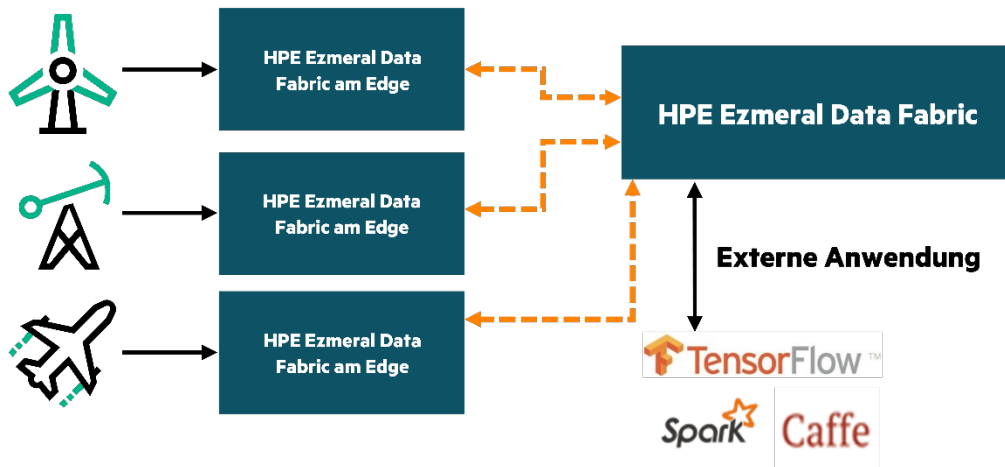


ABBILDUNG 10. HPE Ezmeral Data Fabric erstreckt sich zwischen mehreren Edge-Clustern und Core-Rechenzentren-Clustern

Daten-Volumes in Edge-Clustern erfüllen verschiedene Zwecke und ermöglichen ein fein abgestimmtes Datenmanagement. Beispielsweise können Daten, die auf dem Edge-Cluster analysiert werden sollen, auf einem Volume gespeichert werden, während verarbeitete, extrahierte oder analysierte Daten auf einem anderen Volume gespeichert werden und jederzeit zurück in das Core-Rechenzentrum gespiegelt werden können. Zurück im Core können Daten von zahlreichen Edge-Clustern aggregiert, weiter analysiert oder mit KI-Anwendungen genutzt werden. KI-Modelle können daraufhin auf einem weiteren Volume gespeichert und erneut zum Edge übertragen werden.

Andere Formen der Datenbewegung

Spiegelung ist nicht die einzige Form der Datenbewegung auf Plattformebene. Die HPE Ezmeral Data Fabric unterstützt die Replikation von Data-Fabric-Tabellen und Ereignis-Streams nahezu in Echtzeit.

Tabellenreplikation ist die bevorzugte Art, Daten in Data-Fabric-Tabellen zu bewegen, wobei die Spiegelung von Volumes mit Zeiten ebenfalls funktioniert. Tabellenreplikation ermöglicht eine lokal zeilengenaue Verschiebung von Aktualisierungen zu anderen Tabelleninstanzen, die lokal oder in anderen Clustern liegen können. Diese Replikation kann verkettet werden und Aktualisierungen für viele Tabellen zu einer kombinieren. In Edge-Systemen werden häufig Tabellen zur Speicherung von Datenmanifesten oder Verarbeitungsverläufen für unformatierte Daten und Extrakte verwendet, die separat über Spiegelungen übertragen werden. Tabellenreplikation kann auch genutzt werden, um die Replikation der Daten in kleinen, sehr schnellen Clustern zurück zu einem Core zur langfristigen Aufbewahrung zu ermöglichen. Da die Replikation außerhalb des primären Latenzpfades stattfindet, können diese kleinen schnellen Cluster mit der vollen Geschwindigkeit leistungsstarker Speichergeräte ausgeführt und sogar vollständig im Arbeitsspeicher gehostet werden.

Die Replikation von Ereignis-Streams (Nachrichten-Streams) in der HPE Ezmeral Data Fabric ähnelt der Tabellenreplikation, mit Ausnahme der Einschränkung, dass alle Themen in einem Stream jeweils aus einem einzelnen Ursprungs-Cluster stammen müssen. Dadurch entstehen Architekturen, in denen Streams in vielen Rechenzentren auf einen einzigen globalen Stream repliziert werden, wie in Abbildung 11 illustriert. Diese Streams verwenden Themen, deren Namen ein Präfix des ursprünglichen Rechenzentrums vorangestellt wird, um Kollisionen zwischen Themen zu vermeiden. Solche replizierten Stream-Designs sind äußerst effektiv für die Erfassung von Telemetrieoptionen aus vielen Quellen. Zudem sind sie leicht aufzubauen und zu warten.



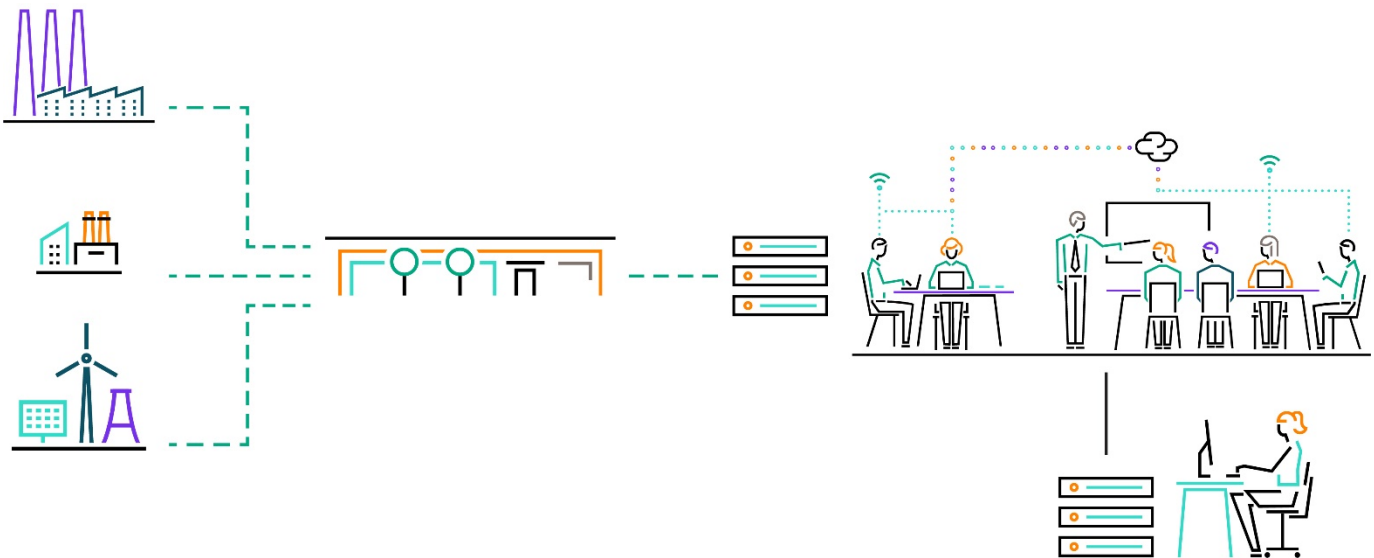


ABBILDUNG 11. Replikation von Ereignis-Streams (dargestellt als grüne gestrichelte Linien) zwischen Edge-Standorten und zentralen Rechenzentren

Die Replikation von Ereignis-Streams ist eine nützliche Möglichkeit, geografisch verteilte Data-Fabric-Cluster miteinander zu verbinden, aber denken Sie daran, dass dies nicht die einzige Möglichkeit ist. Daten in Form von Ereignis-Streams, Tabellen und Dateien können auch durch Spiegelung auf Volume-Ebene zwischen Data-Fabric-Clustern bewegt werden (in Abbildung 11 als orangefarbene gestrichelte Linie dargestellt).

Globale Datenbereitstellung

Die Möglichkeit, die Datenbewegung auf verschiedene Arten auf Plattformebene zu bewältigen, ist ein großer Vorteil. Ebenso vorteilhaft ist es allerdings auch, Daten nicht jedes Mal von Standort zu Standort bewegen zu müssen, wenn Sie darauf zugreifen möchten. Die HPE Ezmeral Data Fabric ermöglicht einen Remote-Datenzugriff in globalen Bereitstellungen dank des globalen Namespace der Data Fabric.

Das bedeutet, dass die verschiedenen HPE Ezmeral Data Fabric Cluster **gegenseitig adressierbar** sind. Dies gilt unabhängig davon, ob sich mehrere Cluster im gleichen Rechenzentrum oder verteilt in verschiedenen Städten befinden. Der Remote-Zugriff funktioniert auch von Core zu Edge oder von Core zu Cloud, unabhängig davon, welcher Cloud-Anbieter eingesetzt wird.

Somit bildet die Data Fabric eine einheitliche Datenebene, die für alle auszuführenden Anwendungen verfügbar ist – ganz gleich, wo sie ausgeführt werden müssen. Analysten, Entwickler, Data Engineers und Data Scientists greifen je nach Bedarf lokal oder remote auf Daten zu. Alternativ können sie mühelos Datenbewegung durch Spiegelung oder die Replikation von Ereignis-Streams und Tabellen einrichten, wie vorher besprochen. Abbildung 12 illustriert diesen globalen Aspekt der Data-Fabric-Bereitstellung.



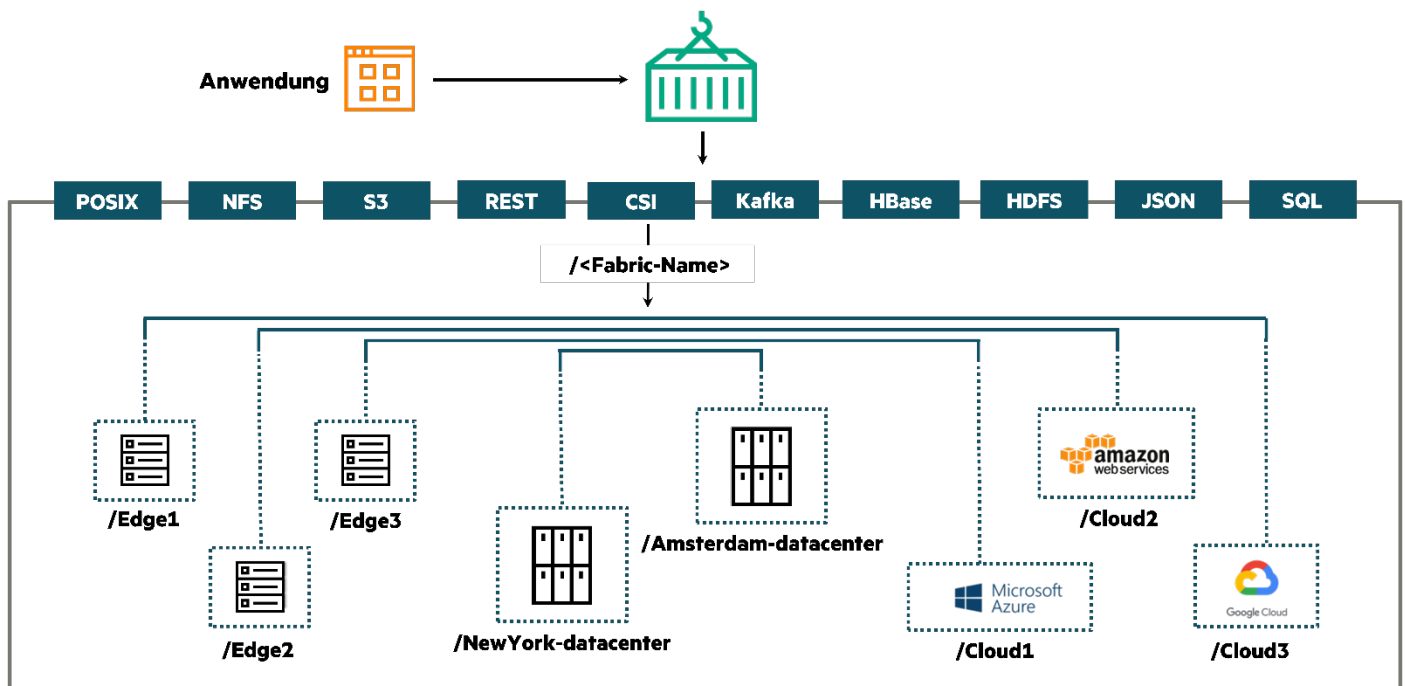


ABBILDUNG 12. HPE Ezmeral Data Fabric für globale Datenbereitstellung

Die Data Fabric ermöglicht einen gegenseitigen Datenzugriff von einem Cluster zum anderen innerhalb des gleichen globalen Namespace. Anwendungen können standortunabhängig auf den richtigen Daten ausgeführt werden.

Beachten Sie die Vielzahl an APIs, die in Abbildung 12 aufgeführt werden. Sie dienen als Beleg für die Flexibilität des von der HPE Ezmeral Data Fabric ermöglichten Datenzugriffs zur Ausführung von Anwendungen, die mit Legacy-Tools sowie mit modernen Analyse- und KI-Tools erstellt wurden. Darüber hinaus geschieht diese gesamte globale Bereitstellung unter der gleichen Administration und der gleichen integrierten Sicherheit, wie im Folgenden erläutert wird.

Die Pfadnamen in der Data Fabric bestehen aus einem Präfix, das den aufzurufenden Cluster angibt, und dem restlichen Pfadnamen mit dem/der zu referenzierenden Verzeichnis, Datei, Tabelle oder Stream. Die Kombination aus Datenbewegung auf Fabric-Ebene und nahtlosem Remote-Zugriff ermöglicht die Zusammenstellung von Systemen mit riesigem Umfang.

Cloud-unabhängige Datenportierbarkeit

Die Daten in einer [HPE Ezmeral Data Fabric](#) können sich auf jeder großen Public Cloud oder auch in einer Private Cloud, in lokalen Rechenzentren oder an Edge-Standorten befinden. Zudem können die Daten an jedem dieser Standorte entweder direkt aufgerufen oder zum Spiegeln oder Replizieren an mehrere Standorte eingerichtet werden, um sie für einen leistungsfähigen lokalen Zugriff zu optimieren. Diese Funktion der Data Fabric bietet einen erheblichen Vorteil, da damit ein System zusammengestellt werden kann, das sich für die aktuellen Anforderungen eignet und gleichzeitig Änderungen zulässt, wenn sich die Anforderungen ändern. Designs von Hybrid Cloud zu On-Premises werden mit der HPE Ezmeral Data Fabric immer beliebter.

Die Möglichkeit, remote auf Daten zuzugreifen und Daten an mehreren Standorten zu platzieren, bedeutet auch, dass Anwendungen je nach geeignetem Hardware- oder Workload-Typ an dem am besten geeigneten Ort ausgeführt werden können. Eine kurze Steigerung der Anzahl Cloud-basierter Maschinen zur vorübergehenden Erweiterung von Rechenressourcen ist in Cloud-Bereitstellungen einfach zu realisieren und kann bei Spitzen-Workloads Kosteneinsparungen ermöglichen. Allerdings ist ein On-Premises-Rechenzentrum bei stabilen, vorhersehbaren Basislasten möglicherweise kosteneffizienter. Die HPE Ezmeral Data Fabric stellt eine Datenebene, die sich zwischen On-Premises- und Cloud-Bereitstellungen und bis zu Edge-Standorten erstreckt, wie in Abbildung 13 illustriert.



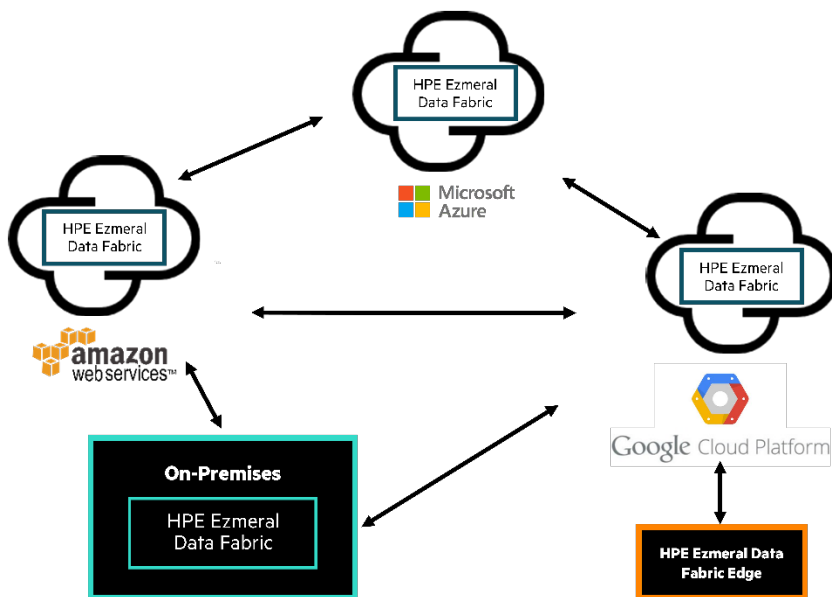


ABBILDUNG 13. Optimierte Ressourcen mithilfe von Hybrid-Cloud-Architekturen

Hybrid-Architekturen ermöglichen eine Optimierung von Ressourcen für stabile statt sporadischen Workloads, wie in Abbildung 13 illustriert. Außerdem macht die Verwendung einer Cloud-unabhängigen Dateninfrastruktur wie der HPE Ezmeral Data Fabric das System unabhängig von einem bestimmten Public-Cloud-Anbieter.

Integriertes Daten-Tiering

Die HPE Ezmeral Data Fabric verfügt über eine Reihe von zusammengehörigen Funktionen, die kollektiv als Daten-Tiering bezeichnet werden und Fabric-Administratoren die Optimierung von Speicherkapazität, Leistung und Kosten ermöglichen. Daten, von denen nicht erwartet wird, dass sie häufig aufgerufen oder aktualisiert werden, können automatisch über verschiedene Optionen der Data Fabric ausgelagert werden.

Transparentes Daten-Tiering über die Data Fabric

HPE Ezmeral Data Fabric unterstützt zwei verschiedene Arten von Daten-Tiering:

- Löschungskodierung innerhalb der Data Fabric
- Daten-Tiering an S3 oder ähnliche Objektspeicher, das automatisch von der Data Fabric gehandhabt wird

Die Nutzung des integrierten Daten-Tierings mit der Data Fabric bietet eine Reihe von Vorteilen gegenüber dem Daten-Tiering über andere Dateninfrastrukturen. Diese Vorteile gelten für beide Optionen (Löschungskodierung und S3-Tiering):

- Das Daten-Tiering wird durch den Benutzer oder durch konfigurierbare Richtlinien aufgerufen, aber automatisch über Prozesse der Data Fabric auf Plattformebene durchgeführt.
- Anwendungen können Daten auch nach dem Tiering weiterhin **nach dem gleichen Pfadnamen** referenzieren.
- Die Schreibleistung wird durch das Daten-Tiering nicht beeinträchtigt, weil die Data Fabric den Prozess (Löschungskodierung oder Tiering auf S3) automatisch im Hintergrund handhabt.
- Im Gegensatz zu vielen anderen Systemen lässt die Data Fabric Sie die Daten aus dem Daten-Tiering erneut aufrufen – es handelt sich um keine Einbahnstraße.

Denken Sie über die Bedeutung des großen Vorteils nach, der in der vorherigen Liste den zweiten Aufzählungspunkt bildet. Benutzer und Anwendungen können die ausgelagerten Daten weiterhin nutzen und bemerken die Änderung nicht einmal, abgesehen von geringfügigen Leistungsunterschieden. Die ausgelagerten Daten können aufgerufen und sogar aktualisiert werden. Werden sie explizit erneut aufgerufen, werden sie in normaler replizierter Form wiederhergestellt (letzter Aufzählungspunkt in der vorherigen Liste). Außerdem können Daten nach Bedarf erneut ausgelagert werden.

Ein weiterer nützlicher Aspekt des Daten-Tierings mit beiden Tiering-Optionen der Data Fabric besteht im einfachen Management und der effizienten Ausführung. **Die Daten-Tiering-Richtlinie wird auf Volume-Ebene festgelegt, aber auf Dateiebene ausgeführt.** Dieses Konzept ist nützlich, weil der Eigentümer des Volumes die gewünschte Richtlinie – z. B. das Tiering einer Datendatei, die in einem bestimmten Zeitraum nicht geändert wurde – zuweisen kann und diese Richtlinie daraufhin als Auslöser für das Volume dient, um nur solche Dateien auszulagern, die das Kriterium erfüllen. Somit kann ein Volume eine Kombination aus ausgelagerten und nicht ausgelagerten Dateien enthalten. Da der Pfadname und die Berechtigungen für ausgelagerte Daten nicht geändert werden, sehen diese für die Anwendung vor und nach dem Tiering gleich aus.



Löschungskodierung mit der HPE Ezmeral Data Fabric

Bei der [Löschungskodierungsoption](#) werden Daten in ein Format kopiert, das weniger Speicherplatz beansprucht, aber das gleiche Maß an Ausfallsicherheit bietet wie eine einfache Replikation. Sehen Sie sich für eine umfassendere Erläuterung der Löschungskodierung das Video „[Praktische Löschungskodierung in einer Data Fabric](#)“ an.

Für Daten-Tiering konfigurierte Volumes bestehen aus zwei Volumes: primäres und Schatten-Volume. Alle Daten, die auf das Volume geschrieben werden, werden auf die primäre Version geschrieben. Wenn die Daten später in eine löschungskodierte Form konvertiert werden, wird diese kodierte Form auf dem Schatten-Volume gespeichert. Auf dem primären Volume verbleiben Stubs zum Referenzieren der Daten auf dem sekundären Volume. Das Schatten-Volume ermöglicht unterschiedliche Speicherort-Einstellungen für die Daten, sobald sie ausgelagert wurden. So können Sie beispielsweise aktive Daten auf Flash-Datenspeicher behalten und auf günstigere Festplattenlaufwerke auslagern, um keinen Platz auf dem teureren Flash-Datenspeicher zu belegen.

Data-Fabric-fähiges Tiering auf Objektspeicher

Mit dem gleichen grundlegenden Mechanismus können Daten auf ein Objektspeichersystem wie AWS S3 oder Azure Blobs ausgelagert werden. Der einzige Unterschied ist, dass das Schatten-Volume durch Objekte auf dem entsprechenden Objektspeicher ersetzt wird. Die Grundidee ist die gleiche: die Verschiebung inaktiver Daten auf eine günstigere Form von Datenspeicher, allerdings ist die Wirkung auf die Leistung viel größer. Wie zuvor können Benutzer weiterhin über die ursprünglichen Pfadnamen auf ihre Daten zugreifen und die Zugriffskontrolle auf den kategorisierten Daten wird auf die gleiche Weise durchgesetzt wie vorher.

Denken Sie daran, dass Daten, die direkt auf S3 oder Azure Blobs gespeichert werden, nicht verändert werden können. Doch überraschenderweise sind Daten, die in der Data Fabric gespeichert und anschließend ausgelagert werden, veränderbar, wodurch diese Methode zu einer bequemerer Möglichkeit wird, von S3 zu profitieren. Darüber hinaus bleiben die Pfadnamen und Zugriffskontrollrichtlinien der ursprünglichen Data-Fabric-Datei in den per Data Fabric ausgelagerten Daten erhalten. Benutzer profitieren von den niedrigeren Kosten von Objektspeicher und höherer Leistung, bis die Daten ausgelagert werden, müssen aber unabhängig von den Einschränkungen von S3 oder Azure Blobs nicht auf den Komfort von stabilen Pfadnamen, mehreren Zugriffs-APIs und veränderbaren Daten verzichten.

Auswirkungen des integrierten Daten-Tierings

Die Transparenz des Daten-Tierings für normale Benutzer ist wichtig, weil Sie den Benutzern damit eine einzige, stabile Ansicht der Daten vorlegen können. Die Entscheidung, Daten auf kostengünstigere Formate oder sogar Sekundärspeicher auszulagern, kann von Administratoren ohne oder nur mit geringen Auswirkungen auf Anwendungen getroffen werden, da alle Dateien im gesamten Prozess unverändert erscheinen.

Das endgültige Resultat sind niedrigere Betriebskosten mit deutlich einfacherem Management. Manchmal wird versucht, ein ähnliches Ergebnis zu erzielen, indem Daten explizit auf S3 verschoben werden. Das ist jedoch problematisch, weil die Umstellung für alle Anwendungen ersichtlich sein muss. Wenn die Daten wieder auf aktiven Datenspeicher abgerufen werden, müssen alle diese Anwendungen erneut geändert werden. Transparentes Daten-Tiering macht die Administration der Data Fabric deutlich einfacher.

Integrierte Sicherheit

Zugriffskontrollen für große Systeme verlangen nach mehr als der Vergabe von Berechtigungen für Dateien und Verzeichnisse wie bei Linux-Dateisystemen. Der Grund dafür ist, dass große Systeme Tausende von Servern, Milliarden oder Billionen von Dateien, Hunderte von Petabytes und Zehntausende Benutzer verwalten müssen. Die HPE Ezmeral Data Fabric wurde speziell für die Handhabung solcher Umgebungen entwickelt.

Wenn Sie von einem einzelnen Linux-Server zu einer Data Fabric wechseln, wird der Datenzugriff auf zwei wesentliche Arten geändert:

- Die Anzahl der zu verwaltenden Objekte nimmt deutlich zu, weil die Data Fabric viel größere Umfänge unterstützt.
- Das Konzept eines Prozessverantwortlichen als Identifikator für einen Linux-Benutzer wird nahezu bedeutungslos, weil Programme jetzt auf Hunderten bis Tausenden von Maschinen ausgeführt werden, häufig mit inkonsistenten Benutzer-IDs.

Um diese Unterschiede zu überwinden, werden Zugriffskontrollen in einer HPE Ezmeral Data Fabric in Form eines ACE ausgedrückt. Die Identität wird mittels eines kryptografischen Zertifikats bestätigt. Bei einem ACE handelt es sich um einen booleschen Ausdruck, der in Bezug auf Benutzer und Gruppen definiert ist. Er gewährt Benutzern die Möglichkeit, ein Verzeichnis oder Objekt zu lesen, zu schreiben, auszuführen, zu scannen oder zu administrieren. Weiter werden ACEs durch die Möglichkeit verbessert, Benutzerrollen zu definieren und einzusetzen. Diese Rollen können den Zweck einer Gruppe erfüllen, setzen aber nicht voraus, dass Gruppen in Ihrem Benutzeridentitäts-Managementsystem definiert werden.

Das Konzept der **Benutzer-ID** hat sich in der Data Fabric ebenfalls gewandelt. Bei einem konventionellen Linux-System hat jeder Prozess effektive und reale Benutzer-IDs. Bei der Data Fabric wird die Identität eines Prozesses durch ein kryptografisches Zertifikat im Besitz des Prozesses definiert und an einem bekannten Ort gespeichert. Dieses Zertifikat definiert die Identität, die in ACEs verwendet werden soll, mit denen der Zugriff auf Objekte in der Data Fabric kontrolliert wird. Um dem Großteil der Benutzer das Leben zu erleichtern, wird bei der Anmeldung ein Zertifikat erstellt. Im Gegensatz dazu erhält bei Containern, die von Kubernetes verwaltet werden, jeder Container Zugriff auf ein Ticket, das in einem Kubernetes-Geheimsschlüssel gespeichert ist, ohne dass eine Aktion des Containers erforderlich ist.

Im Allgemeinen behält das [Sicherheitssystem der Data Fabric](#) die besten Aspekte der Genehmigungskonzepte von Linux bei, weitet sie aber für das neue Ökosystem umfangreicher Anwendungen aus, die auf vielen Maschinen ausgeführt werden und auf riesige Datenmengen zugreifen.

Richtlinienbasierte Sicherheit

Die Daten-Governance wird durch die richtlinienbasierte Sicherheit der HPE Ezmeral Data Fabric weiter verbessert, die die Kontrolle über den Datenzugriff skaliert. Dieses Merkmal bedeutet, dass mit der Data Fabric weniger menschliche Administratoren viel größere Datenmengen verwalten können. So können die Daten skaliert werden, ohne die Größe des IT-Teams zu skalieren.



Bei richtlinienbasierter Sicherheit wird eine Richtlinie auf Datenobjekte in der Data Fabric angewendet, allerdings wird die Richtlinie an einer anderen Stelle in der Data Fabric definiert. Da der Inhalt der Richtlinie nicht direkt an die Dateien und Verzeichnisse angehängt wird, ist es möglich, die Richtlinie später global zu ändern, ohne die Datenobjekte einzeln adressieren zu müssen. Die Richtlinienänderung tritt zentral auf, wird aber auf alle Objekte angewendet, die die Richtlinie referenzieren. Das System setzt die Zugriffskontrolle dementsprechend auf alle Daten durch, ohne die Genehmigungen für Millionen, Milliarden oder gar Billionen von Dateien aktualisieren zu müssen, wie in Abbildung 14 illustriert.

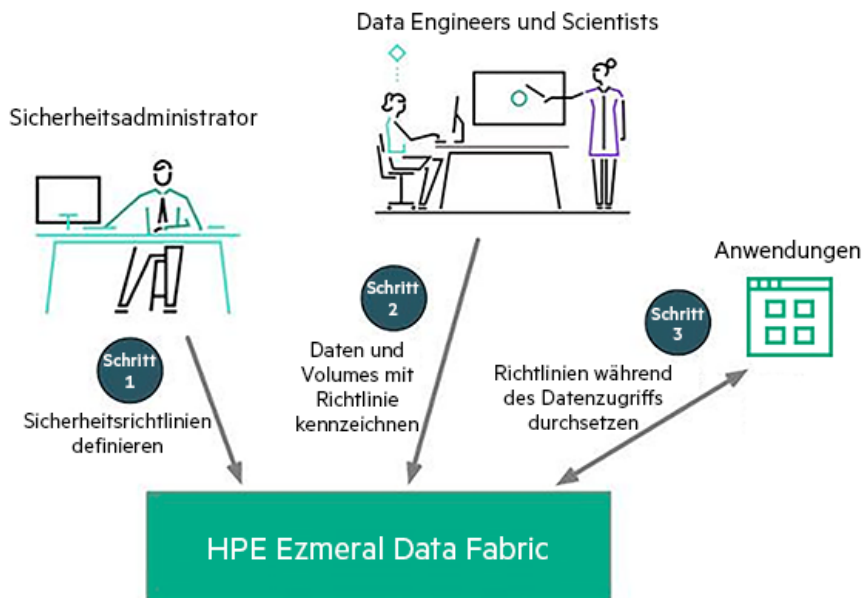


ABBILDUNG 14. Richtlinienbasierte Sicherheit in der HPE Ezmeral Data Fabric

Die richtlinienbasierte Sicherheit der Data Fabric macht die Anwendung oder Anpassung einer einheitlichen Sicherheit auf Plattformebene einfach und effizient, selbst bei umfangreichen Systemen mit vielen Benutzern, Petabytes an Daten und Millionen von Dateien. Richtlinienbasierte Sicherheit ist ein enormer Vorteil für umfangreiche Datensysteme, weil es möglich ist, die Verfügbarkeit von Millionen von Objekten gleichzeitig zu ändern, aber auch weil das Recht, eine Richtlinie anzuwenden, nicht das Gleiche ist wie das Recht, sie zu ändern.

HPE EZMERAL DATA FABRIC DATEISYSTEM

Wer ein hochgradig skalierbares verteiltes Datenspeicherungs- und -verwaltungssystem erwartet, das nur Dateien nutzt, bekommt von HPE als Option eine eigenständige Version des verteilten Dateisystems der HPE Ezmeral Data Fabric. Diese Option bietet die in diesem Dokument bereits beschriebenen Funktionen für Dateien, einschließlich des flexiblen Datenzugriffs über verschiedene APIs, des globalen Namespace, des Datenmanagements über Data-Fabric-Volumes, effizienter Datenbewegung durch Spiegelung, extremer Skalierbarkeit, Leistung, Zuverlässigkeit und Ausfallsicherheit – das alles mit integrierter Sicherheit.

Die rein dateibasierte Option der HPE Ezmeral Data Fabric beinhaltet nicht die integrierten Tabellen und Streams und der Zugriff auf Daten in der Data Fabric über die S3-API wird nicht unterstützt. Daten-Tiering wird unterstützt, einschließlich Löschungskodierung und Tiering von der Data Fabric auf S3-Objektspeicher.

Der CSI-Treiber ist ebenfalls enthalten und ermöglicht containerisierten Anwendungen, die mit Kubernetes orchestriert werden, den Zugriff auf Dateien in der Data Fabric sowie den Erhalt der Datenausgabe aus zustandsabhängigen Anwendungen in der Data Fabric.

NUTZUNG DER HPE EZMERAL DATA FABRIC

Die Nutzung der HPE Ezmeral Data Fabric als primäre Dateninfrastruktur in umfangreichen Systemen vereinfacht die Architektur und Workflows für Analysen und KI in zahlreichen Anwendungsfällen. Abbildung 15 illustriert die wichtigsten Funktionen der HPE Ezmeral Data Fabric als einheitliche Datenebene, durch die die in diesem Whitepaper besprochenen Vorteile ermöglicht werden.



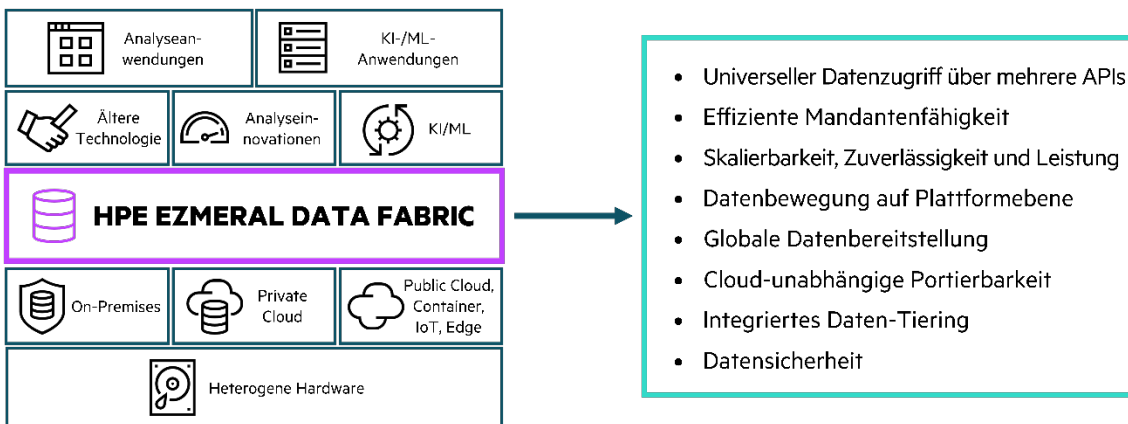


ABBILDUNG 15. Wichtigste Funktionen der HPE Ezmeral Data Fabric

Die HPE Ezmeral Data Fabric ist branchenweit einzigartig als hochgradig skalierbare und ausfallsichere Dateninfrastruktur, die hervorragende Leistung für global verteilte Systeme einschließlich Edge- und Hybrid-Cloud-Architekturen bietet. Globale Sicherheit und Administrationsrichtlinien gewährleisten eine sichere und konsistente gemeinsame Datennutzung, wodurch Mandantenfähigkeit praktikabel und wünschenswert wird.

Die unmittelbaren Vorteile bei der Verwendung der Data Fabric liegen auf der Hand, wie in diesem Dokument beschrieben. Aber die HPE Ezmeral Data Fabric erfüllt nicht nur die Kriterien für eine unternehmensweite vereinheitlichende Dateninfrastruktur, sondern berücksichtigt auch die Zukunft. Systeme mit der Data Fabric als allgemeine Datenebene können zu viel größeren Umfängen ausgebaut oder auf neue Standorte bzw. um neue Anwendungen und Projekte erweitert werden, **ohne sie umgestalten oder ein separates System zusammenstellen zu müssen**. Zudem können diese Änderungen – darunter auch vorübergehende, saisonale oder langfristige Erweiterungen des Umfangs – effizient vorgenommen werden, ohne IT-Ressourcen skalieren zu müssen.

Die HPE Ezmeral Data Fabric reduziert heute und in Zukunft die Kosten und befreit Entwickler, Data Scientists, Analysten und IT-Experten von mühseligen Infrastrukturaufwänden, damit sie sich auf ihre eigenen Aufgaben konzentrieren können. Das Ergebnis: höhere Effizienz, schnellere Entwicklung, zuverlässige Produktion und Offenheit für Innovationen.

RESSOURCEN

- [HPE Ezmeral Data Fabric Webseite](#)
- [HPE Ezmeral Data Fabric Dokumentation](#)
- [HPE Ezmeral Data Fabric Plattformseite in der HPE Developer Community](#)
- [Schnellere Entwicklung autonomer Fahrzeuge durch einfachen Zugriff auf eine globale Data Fabric](#)
- [Schnellere Erkenntnisse aus Daten für ein besseres Arbeitsleben](#)
- [„Stretching HPE Ezmeral Data Fabric from Edge-To-Cloud“](#), Data Fabric Edge Webinar, Jimmy Bates, Chad Smykay, Fabian Wilckens
- [Lightboard-Video „Einheitliche Data Fabric vom Edge über den Core bis zur Cloud“](#) mit Jimmy Bates
- [Lightboard-Video „Lösung für große IoT-Probleme“](#) mit Ted Dunning
- [Kostenloses E-Book KI und Analysen im großen Maßstab: Lektionen aus echten Produktionssystemen](#) von Ted Dunning und Ellen Friedman herunterladen
- [Analystenbericht – IDC-Whitepaper – Nutzen Sie das Potenzial: der Business Value von HPE Ezmeral Data Fabric](#) herunterladen



Entscheiden Sie sich für das richtige Produkt.
Kontaktieren Sie unsere Presales-Experten.



Chat



E-Mail



Telefon



Updates abrufen