DELL Technologies

White paper tecnico

Panoramica e architettura di ECS

Abstract

Questo documento offre una panoramica tecnica e informazioni sulla progettazione della piattaforma di object storage software-defined su scala cloud Dell EMC™ ECS™.

Febbraio 2021

H14071.18

Revisioni

Data	Descrizione
Dicembre 2015	Release iniziale
Maggio 2016	Aggiornato per 2.2.1
Settembre 2016	Aggiornato per 3.0
Agosto 2017	Aggiornato per 3.1
Marzo 2018	Aggiornato per 3.2
Settembre 2018	Aggiornato per l'hardware Gen3
Febbraio 2019	Aggiornato per 3.3
Settembre 2019	Aggiornato per 3.4
Febbraio 2020	Modifiche aggiornate a ECSDOC-628
Maggio 2020	Aggiornato per 3.5
Novembre 2020	Aggiornato per 3.6
Febbraio 2021	Aggiornato per 3.6.1

Ringraziamenti

Questo documento è stato prodotto da quanto segue:

Autore: Zhu, Jarvis

Sommario

Executive Summary

Le organizzazioni richiedono opzioni per l'utilizzo di servizi cloud pubblici con l'affidabilità e il controllo di un'infrastruttura cloud privata. Dell EMC ECS è una piattaforma di object storage software-defined, supportata da IPv6 e su scala cloud, che offre servizi di storage S3, Atmos, CAS, Swift, NFSv3 e HDFS su un'unica piattaforma moderna.

Grazie a ECS, gli amministratori possono gestire facilmente un'infrastruttura di storage distribuita a livello globale in un unico namespace che fornisce accesso al contenuto da qualsiasi posizione. I componenti core di ECS sono a più livelli per garantire flessibilità e resilienza. Ogni livello è astratto e scalabile in modo indipendente e offre high availability.

Gli sviluppatori adottano un semplice accesso all'API RESTful per i servizi di storage. L'uso della semantica HTTP, come GET e PUT, semplifica la logica dell'applicazione richiesta rispetto alle familiari e tradizionali operazioni di file basate sui percorsi. Inoltre, il sistema di storage sottostante di ECS è estremamente coerente, il che significa che può garantire una risposta autorevole. Le applicazioni richieste per garantire una distribuzione dei dati basata sulla autorizzazioni sono in grado di farlo senza logica di programmazione complessa, utilizzando semplicemente ECS.

1 Introduzione

Questo documento fornisce una panoramica della piattaforma di object storage Dell EMC ECS. Descrive in dettaglio l'architettura di progettazione e i componenti core di ECS, ad esempio i servizi di storage e i meccanismi di protezione dei dati.

1.1 Audience

Questo documento è destinato a chiunque sia interessato a comprendere il valore e l'architettura di ECS. Mira a fornire il contesto con link a informazioni aggiuntive.

1.2 Ambito

Questo documento è incentrato principalmente sull'architettura ECS. Non copre le procedure di installazione, amministrazione e upgrade del software o hardware ECS. Inoltre, non copre le specifiche sull'utilizzo e la creazione di applicazioni con le API di ECS.

Gli aggiornamenti a questo documento vengono eseguiti periodicamente e in genere coincidono con le versioni principali o con le nuove funzioni.

2 Valore di ECS

ECS offre un valore significativo alle aziende e ai fornitori di servizi alla ricerca di una piattaforma progettata per supportare una rapida crescita dei dati. I principali vantaggi e funzioni di ECS che consentono alle aziende di gestire e archiviare globalmente i contenuti distribuiti su larga scala includono:

- **Scala cloud:** ECS è una piattaforma di object-storage per carichi di lavoro tradizionali e di nuova generazione. L'architettura a software-defined flessibile e a più livelli promuove una scalabilità illimitata. Caratteristiche principali:
  - Infrastruttura a object distribuita a livello globale
  - Scalabilità Exabyte+ senza limiti di capacità del pool di storage, del cluster o dell'ambiente federato
  - Non esistono limiti al numero di object in un sistema, namespace o bucket
  - Efficiente con i carichi di lavoro contenenti file sia di piccole che di grandi dimensioni senza limiti in termini dimensioni degli object
- **Deployment flessibile:** ECS ha una flessibilità senza pari con funzionalità quali:
  - Deployment dell'appliance
  - Deployment solo software con supporto di hardware standard del settore certificato o personalizzato
  - Supporto multiprotocollo: Object (S3, Swift, Atmos, CAS) e file (HDFS, NFSv3)
  - Più carichi di lavoro: app moderne e archiviazione a lungo termine
  - Storage secondario per Data Domain Cloud Tier e Isilon tramite CloudPools
  - Percorsi di aggiornamento senza interruzioni per i modelli ECS della generazione attuale
- **Livello enterprise:** ECS offre ai clienti un maggiore controllo dei data asset con storage di livello enterprise in un sistema sicuro e conforme con funzioni quali:
  - Dati at-rest (DARE) con rotazione delle chiavi e gestione delle chiavi esterne.
  - Comunicazione tra siti crittografata
  - Disabilita le porte 9101/9206 per impostazione predefinita, consentendo alle organizzazioni di soddisfare le policy di conformità
  - Reporting, retention dei record basata su policy ed eventi e protezione avanzata della piattaforma per la conformità alla regola SEC 17a-4(f), inclusa la gestione avanzata della retention, come, ad esempio, controversie e governance min-max
  - Conformità con le linee guida di protezione avanzata STIG (Security Technical Implementation Guide) di DISA (Defense Information Systems Agency)
  - Controlli di autenticazione, autorizzazione e accesso con Active Directory e LDAP
  - Integrazione con l'infrastruttura di monitoraggio e avvisi (trap SNMP e SYSLOG)
  - Funzionalità aziendali migliorate (multi-tenancy, monitoraggio della capacità e avvisi)
- **Riduzione del TCO:** ECS può ridurre drasticamente il costo totale di proprietà (TCO) rispetto allo storage tradizionale e al cloud storage pubblico. Offre anche un TCO inferiore rispetto al nastro per la retention a lungo termine. Ecco alcune delle funzionalità offerte:
  - Namespace globale
  - Prestazioni di file di piccole e grandi dimensioni
  - Migrazione di Centera senza soluzione di continuità
  - Conformità totale ad Atmos REST
  - Overhead di gestione ridotto
  - Ingombro ridotto del data center
  - Utilizzo dello storage elevato

La progettazione di ECS è ottimizzata per i seguenti casi d'uso principali:

- **Applicazioni moderne:** ECS è progettato per lo sviluppo moderno, ad esempio per applicazioni web, mobili e cloud di nuova generazione. Lo sviluppo delle applicazioni è semplificato con una storage estremamente coerente. Insieme all'accesso in lettura e scrittura multiutente simultaneo e multisito, man mano che la capacità ECS cambia e cresce, gli sviluppatori non devono mai ricodificare le proprie app.
- **Storage secondario:** ECS viene utilizzato come storage secondario per liberare lo storage primario dei dati a cui si accede raramente, mantenendolo al contempo ragionevolmente accessibile. Ne sono esempi i prodotti di tiering basati su policy, ad esempio Data Domain Cloud Tier e Isilon CloudPools. GeoDrive, un'applicazione basata su Windows, offre ai sistemi Windows accesso diretto a ECS per archiviare i dati.
- **Archivio con protezione geografica:** ECS può fungere da cloud on-premise sicuro e conveniente a scopo di archiviazione e retention a lungo termine. L'utilizzo di ECS come tier di archiviazione può ridurre in modo significativo le capacità dello storage primario. Oltre al valore predefinito di 12+4, per i casi d'uso di archiviazione a freddo utilizzare uno schema di codifica di erasure (EC) di 10+2 per consentire uno storage più efficiente.
- **Repository di contenuti globali:** i repository di contenuti non strutturati contenenti dati come immagini e video sono spesso archiviati in sistemi di storage molto costosi, rendendo impossibile per le aziende gestire una forte crescita dei dati a costi contenuti. ECS consente di consolidare più sistemi di storage in un unico repository di contenuti, efficiente e accessibile globalmente.
- **Storage per Internet of Things:** l'Internet of Things (IoT) offre una nuova opportunità di entrate per le aziende che possono ricavare valore dai dati dei clienti. ECS offre un'architettura IoT efficiente per la raccolta di dati non strutturati su vastissima scala. Senza limiti sul numero di object, sulle dimensioni degli object o sui metadati personalizzati, ECS è la piattaforma ideale per archiviare i dati IoT. ECS può anche semplificare alcuni flussi di lavoro di analisi: consente infatti di analizzare dati direttamente sulla piattaforma ECS senza richiedere lunghi processi di estrazione, trasformazione e caricamento (ETL). I cluster Hadoop possono eseguire query utilizzando i dati archiviati in ECS da un'altra API di protocollo, ad esempio S3 o NFS.
- **Repository delle prove dei sistemi di videosorveglianza:** a differenza dei dati IoT, i dati di videosorveglianza hanno un numero di object storage molto più basso, ma un ingombro di capacità per file molto più elevato. Sebbene l'autenticità dei dati sia importante, la conservazione dei dati non è così critica. ECS può essere un'area di destinazione a basso costo o una posizione di storage secondario per tali dati. Il software di gestione video può sfruttare le ricche funzionalità personalizzate dei metadati per l'etichettatura dei file con dettagli importanti come la posizione della fotocamera, i requisiti di retention e i requisiti di protezione dei dati. Inoltre, i metadati possono essere utilizzati per impostare il file sullo stato read-only per garantire una catena di custodia sul file.
- **Data lake e analisi:** i dati e l'analisi sono diventati un fattore di differenziazione concorrenziale e una fonte primaria di valore per le organizzazioni. Tuttavia, trasformare i dati in un prezioso asset aziendale è un argomento complesso che può facilmente comportare l'utilizzo di dozzine di tecnologie, strumenti e ambienti. ECS fornisce una serie di servizi per aiutare i clienti a raccogliere, archiviare, gestire e analizzare i dati su qualsiasi scala.

3 dell'architettura

ECS è progettato con alcuni principi di progettazione core, come il namespace globale con un'estrema coerenza; capacità di scalabilità orizzontale, multi-tenancy sicura; prestazioni superiori per object sia di piccole che di grandi dimensioni. ECS è realizzato come sistema completamente distribuito seguendo il principio delle applicazioni cloud, in cui ogni funzione nel sistema è realizzata come un livello indipendente. Con questa progettazione, ogni livello è scalabile in orizzontale su tutti i nodi del sistema. Le risorse vengono distribuite fra tutti i nodi per aumentare la disponibilità e condividere il carico.

In questa sezione viene approfondita l'architettura di ECS e la progettazione del software e dell'hardware.

3.1 Panoramica

ECS viene implementato su un set di hardware standard del settore qualificato o come appliance di storage pronto all'uso. I componenti principali di ECS sono i seguenti:

- **Portale ECS e servizi di provisioning:** WebUI e CLI basati su API per self-service, automazione, reporting e gestione dei nodi ECS. Questo livello gestisce anche servizi di licenze, autenticazione, multi-tenancy e provisioning, ad esempio la creazione di namespace.
- **Data service:** servizi, strumenti e API per supportare l'accesso di object e file al sistema.
- **Engine di storage:** servizio core responsabile dell'archiviazione e del recupero dei dati, della gestione delle transazioni e della protezione e replica dei dati a livello locale e tra siti.
- **Fabric:** servizio di clustering per la gestione e la generazione di avvisi correlati a integrità, configurazione e upgrade.
- **Infrastruttura:** SUSE Linux Enterprise Server 12 per il sistema operativo di base nell'appliance pronto all'uso o nei sistemi operativi Linux qualificati per la configurazione hardware standard del settore.
- **Hardware:** un appliance pronto all'uso o hardware standard del settore qualificato.

La Figura 1 mostra una rappresentazione grafica di questi livelli, che sono descritti in dettaglio nelle sezioni che seguono.

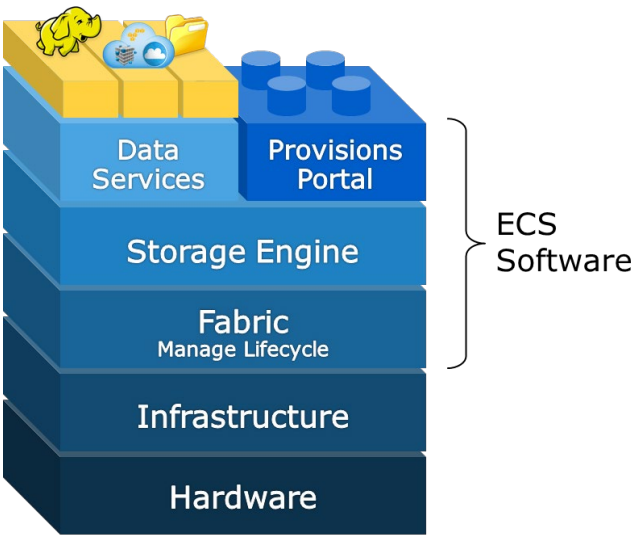

Figura 1 Livelli dell'architettura ECS

3.2 Portale ECS e servizi di provisioning

Gli Storage Administrator gestiscono ECS utilizzando il portale ECS e i servizi di provisioning. ECS fornisce una GUI web-based (WebUI) per gestire, concedere in licenza ed eseguire il provisioning dei nodi ECS. Il portale dispone di funzionalità di reporting complete che includono:

- Utilizzo della capacità per sito, pool di storage, nodo e disco.
- Monitoraggio delle prestazioni in termini di latenza, throughput e avanzamento della replica.
- Informazioni di diagnostica, ad esempio lo stato di ripristino del nodo e del disco.

Il dashboard ECS fornisce informazioni generali sull'integrità e sulle prestazioni a livello di sistema. Questa vista unificata migliora la visibilità complessiva del sistema. Gli avvisi notificano agli utenti gli eventi critici, ad esempio limiti di capacità, limiti di quota, errori del disco o del nodo o gli errori software. ECS fornisce anche un'interfaccia della riga di comando per installare, effettuare l'upgrade e monitorare ECS. L'accesso ai nodi per l'uso della riga di comando viene eseguito tramite SSH. La Figura 2 che segue mostra uno screenshot del dashboard di ECS.

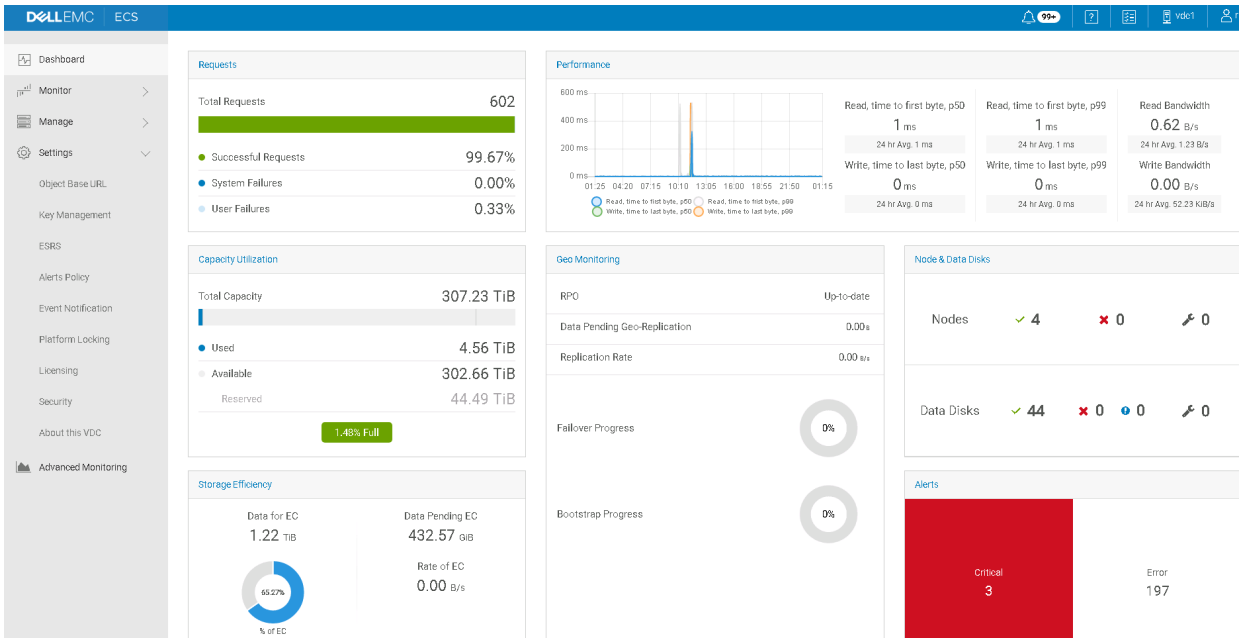

Figura 2 Dashboard dell'interfaccia utente web di ECS

Il reporting dettagliato sulle prestazioni è disponibile nell'interfaccia utente nella cartella Advance Monitoring. I report vengono visualizzati in un dashboard Grafana. Sono disponibili filtri per eseguire il drill-down di namespace, protocolli o nodi specificati. Un esempio di report sulle prestazioni del protocollo S3 è riportato di seguito nella Figura 3.

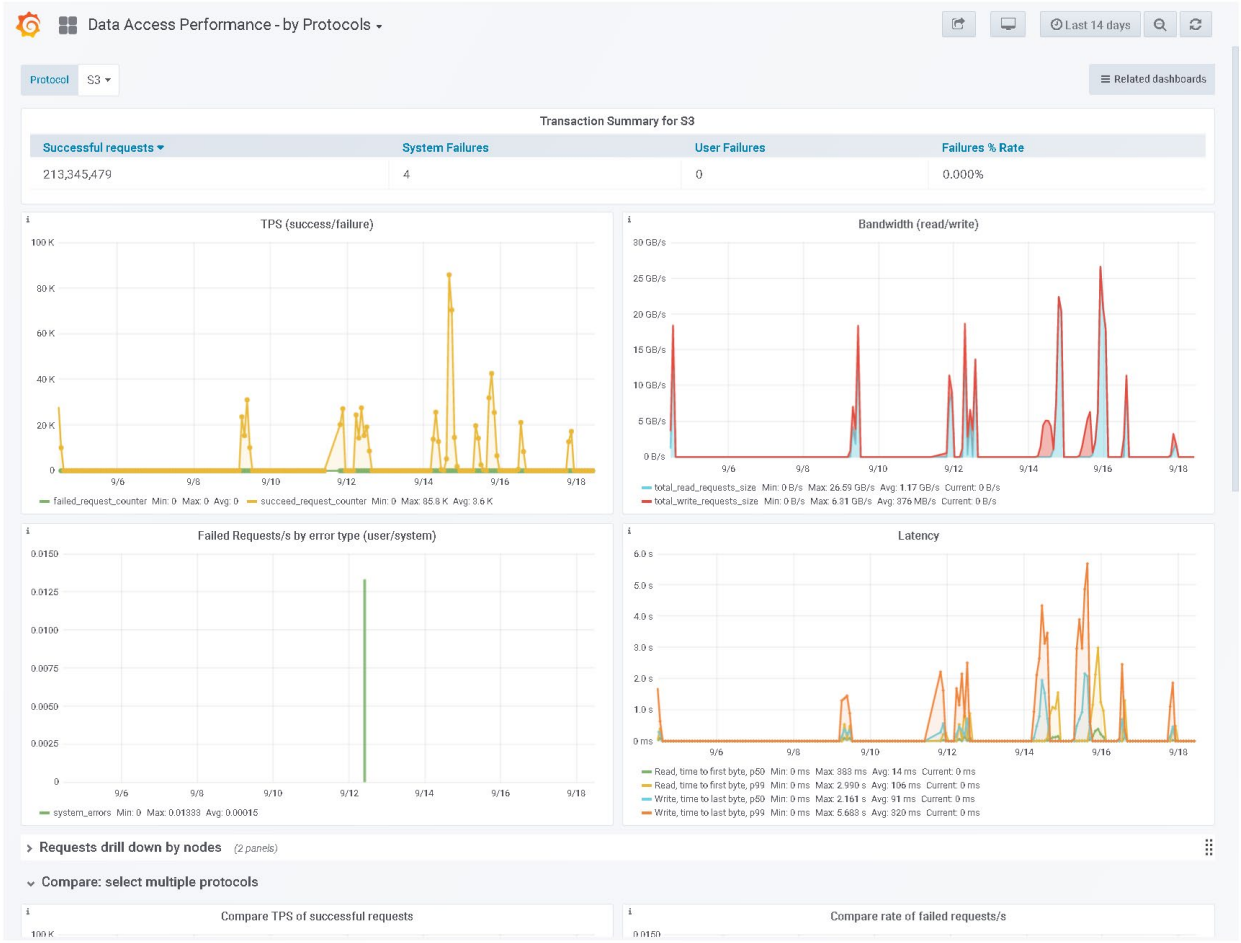

Figura 3 Visualizzazione di monitoraggio avanzato con Grafana

ECS può essere gestito anche utilizzando le API RESTful. L'API di gestione consente agli utenti di amministrare ECS all'interno dei propri strumenti, script e applicazioni nuove o esistenti. L'interfaccia utente web e gli strumenti da riga di comando di ECS vengono creati utilizzando le API di gestione REST di ECS.

ECS supporta i seguenti server di notifica degli eventi che possono essere impostati tramite l'interfaccia utente web, l'API o la CLI:

- Server SNMP (Simple Network Management Protocol)
- Server Syslog

La *Guida dell'amministratore di ECS* contiene ulteriori informazioni e dettagli sulla configurazione dei servizi di notifica.

3.3 Data service

I metodi object e file standard vengono utilizzati per accedere ai servizi di storage di ECS. Nel caso di S3, Atmos e Swift, per l'accesso vengono utilizzate API RESTful su HTTP. Per Content Addressable Storage (CAS), viene utilizzato un metodo di accesso proprietario/SDK. ECS supporta in modo nativo tutte le procedure NFSv3 ad eccezione di LINK. È ora possibile accedere ai bucket ECS tramite S3a.

ECS fornisce l'accesso multiprotocollo in cui è possibile accedere ai dati acquisiti mediante un protocollo tramite altri. Ciò significa che i dati possono essere acquisiti tramite S3 e modificati tramite NFSv3 o Swift, o viceversa. Esistono alcune eccezioni all'accesso multiprotocollo a causa della semantica del protocollo e delle rappresentazioni della progettazione del protocollo. La Tabella 1 mette in evidenza i metodi di accesso e i protocolli che interagiscono.

Tabella 1 Interoperabilità dei data service e protocolli supportati da ECS

Protocolli		**Supportato**	**Interoperabilità**
Object	S3	Funzionalità aggiuntive come Byte Range Updates e Rich ACLS	HDFS, NFS, Swift
	Atmos	Versione 2.0	NFS (solo object basati su percorsi e non basati sullo stile ID object)
	Swift	API V2 e autenticazione Swift e Keystone v3	HDFS, NFS, S3
	CAS	SDK v3.1.544 o versione successiva	N/D
File	HDFS	Compatibilità con Hadoop 2.7	S3, NFS, Swift
	NFS	NFSv3	S3, Swift, HDFS, Atmos (solo object basati su percorsi e non basati sullo stile ID object)

I data service, detti anche head service, sono responsabili della ricezione delle richieste dei client, dell'estrazione delle informazioni necessarie e del loro trasferimento all'engine di storage per ulteriori elaborazioni. Tutti gli head service sono combinati in un unico processo, *dataheadsvc*, in esecuzione all'interno del livello dell'infrastruttura. Questo processo viene ulteriormente incapsulato all'interno di un container Docker denominato *object-main* che viene eseguito su tutti i nodi in ECS. La sezione *Infrastruttura* di questo documento illustra Docker in modo più dettagliato. I requisiti delle porte del servizio protocollo ECS, ad esempio la porta 9020 per la comunicazione S3, sono disponibili nella più recente *Guida alla configurazione della sicurezza di ECS*.

3.3.1 Object

ECS supporta API S3, Atmos, Swift E CAS per l'accesso agli object. Ad eccezione di CAS, gli object o i dati vengono scritti, recuperati, aggiornati ed eliminati tramite chiamate HTTP o HTTPS di GET, POST, PUT, DELETE e HEAD. Per CAS, vengono utilizzati la comunicazione TCP standard e chiamate e metodi di accesso specifici.

ECS fornisce una struttura per la ricerca di metadati per gli object che utilizzano un linguaggio di query avanzato. Si tratta di una potente funzione di ECS che consente ai client di object S3 di cercare object all'interno dei bucket utilizzando il sistema e metadati personalizzati. Sebbene sia possibile effettuare la ricerca utilizzando qualsiasi metadato nei metadati configurati in modo specifico per l'indicizzazione in un bucket, ECS può restituire le query più rapidamente, in particolare per i bucket con miliardi di object.

È possibile indicizzare fino a trenta campi di metadati definiti dall'utente per bucket. I metadati vengono specificati al momento della creazione del bucket. La funzione di ricerca dei metadati può essere abilitata sui bucket con la crittografia lato server abilitata; tuttavia, gli attributi di metadati utente indicizzati utilizzati come chiave di ricerca non verranno crittografati.

Nota: si verifica un impatto sulle prestazioni quando si scrivono dati in bucket configurati per l'indicizzazione dei metadati. L'impatto sulle operazioni aumenta con l'aumentare del numero di campi indicizzati. L'impatto sulle prestazioni richiede un'attenta considerazione sulla scelta se indicizzare i metadati in un bucket e, in caso affermativo, sul numero di indici da gestire.

Per gli object CAS, l'API di query CAS offre una capacità simile per cercare gli object in base ai metadati gestiti per gli object CAS che non devono essere abilitati in modo esplicito.

Per ulteriori informazioni sulle API di ECS e le API per i metadati di ricerca, consultare la *Guida di accesso ai dati ECS* più recente. Per gli SDK di Atmos e S3, visitare il sito GitHub Dell EMC Data Services SDK oppure Dell EMC ECS. Per CAS, visitare il sito della community di Centera. L'accesso a numerosi esempi, risorse e assistenza per gli sviluppatori è disponibile nella community di ECS.

Le applicazioni client, come S3 Browser e Cyberduck, consentono di testare o accedere rapidamente ai dati memorizzati in ECS. ECS Test Drive viene fornito gratuitamente da Dell EMC; consente l'accesso a un sistema ECS pubblico per scopi di test e sviluppo. Dopo la registrazione a ECS Test Drive, agli endpoint REST vengono fornite le credenziali utente per ognuno dei protocolli di object. Chiunque può utilizzare ECS Test Drive per testare l'applicazione API S3.

Nota: solo il numero di metadati che possono essere indicizzati per bucket è limitato a trenta in ECS. Non esiste alcuna limitazione al numero totale di metadati personalizzati archiviati per object, ma solo al numero indicizzato per la ricerca rapida.

3.3.2 HDFS

ECS è in grado di archiviare i dati dei file system Hadoop. In qualità di file system compatibile con Hadoop, le organizzazioni possono creare repository di Big Data su ECS che l'analisi Hadoop può utilizzare ed elaborare. Il data service HDFS è compatibile con Apache Hadoop 2.7, con il supporto per ACL dettagliate e attributo di file system esteso.

ECS è stato convalidato e testato con Hortonworks (HDP 2.7). ECS offre anche supporto per servizi come YARN, MapReduce, Pig, Hive/Hiveserver2, HBase, Zookeeper, Flume, Spark e Sqoop.

3.3.2.1 Supporto per Hadoop S3A

ECS supporta il client Hadoop S3A per l'archiviazione dei dati Hadoop. S3A è un connettore open source per Hadoop basato sull'SDK ufficiale di Amazon Web Services (AWS). È stato creato per risolvere i problemi di dimensionamento e costi dello storage che molti amministratori Hadoop avevano con HDFS. Hadoop S3A connette i cluster Hadoop a qualsiasi object store compatibile con S3, nel cloud pubblico, ibrido oppure on-premise.

Nota: il supporto S3A è disponibile su Hadoop 2.7 o versione successiva

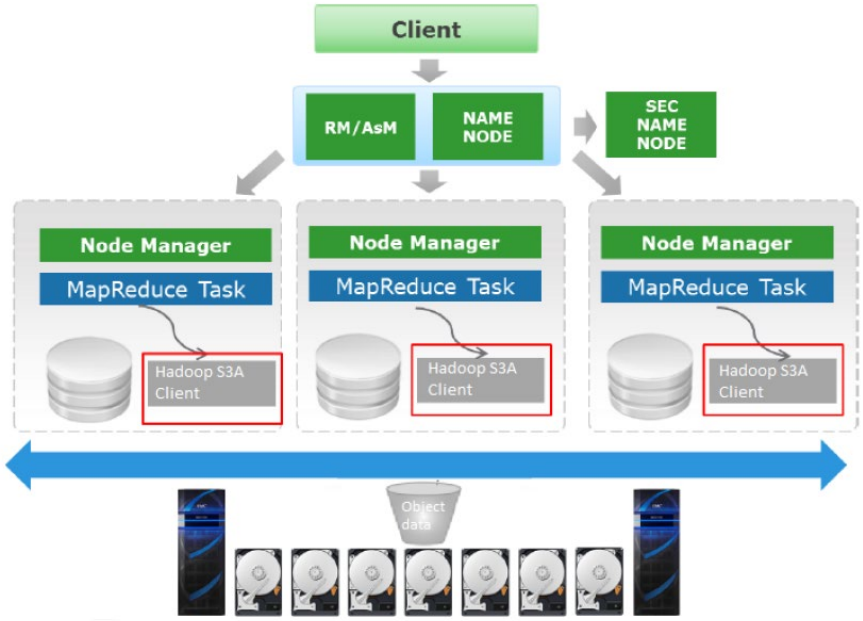

Figura 4 Architettura Hadoop ed ECS

Come illustrato nella Figura 4, quando il cliente configura il cluster Hadoop su HDFS tradizionale, la sua configurazione S3A punta ai dati degli object ECS per eseguire tutte le attività HDFS. Su ogni nodo HDFS Hadoop, qualsiasi componente Hadoop tradizionale utilizzerebbe il client S3A di Hadoop per eseguire l'attività HDFS.

Analisi della configurazione di Hadoop tramite la ECS Service Console

La ECS Service Console (SC) può leggere e interpretare i parametri di configurazione di Hadoop in relazione alle connessioni a ECS per S3A. Inoltre, la SC fornisce una funzione, *Get_Hadoop_Config*, che legge la configurazione del cluster Hadoop e verifica i valori delle impostazioni S3A per individuare errori di scrittura e altri errori. Contattare il team di supporto ECS per assistenza con l'installazione della ECS SC.

Implementazione di Privacera con Hadoop S3A

Privacera è un fornitore terzo che ha implementato un agent lato client Hadoop e l'integrazione con Ambari per la sicurezza granulare di S3 (AWS ed ECS). Sebbene Privacera supporti Cloudera Distribution of Hadoop (CDH), Cloudera (altro fornitore terzo) non supporta Privacera su CDH.

Nota: gli utenti di CDH devono utilizzare i servizi di sicurezza ECS IAM. Se si vuole un accesso protetto a S3A senza utilizzare ECS IAM, contattare il team di supporto.

Per ulteriori informazioni sul supporto S3A, vedere la *Guida all'accesso ai dati ECS* più recente

Sicurezza di Hadoop S3A

ECS IAM consente all'amministratore di Hadoop di configurare policy di accesso per controllare l'accesso ai dati Hadoop S3A. Una volta definite le policy di accesso, sono disponibili due opzioni di accesso utente che gli amministratori di Hadoop possono configurare:

- Utenti/gruppi IAM
  - Creazione di gruppi IAM associati alle policy
  - Creazione di utenti IAM che sono membri di un gruppo IAM

- Asserzioni SAML (utenti federati)
  - Creazione di ruoli IAM associati alle policy
  - Configurazione di CrossTrustRelationship tra Identity Provider (AD FS) ed ECS che esegue il mapping dei gruppi AD ai ruoli IAM

L'amministratore di ECS e l'amministratore di Hadoop devono collaborare per predefinire le policy appropriate. Gli esempi illustrati di seguito descrivono tre tipi di utenti di Hadoop per i quali verranno create le policy. Questi sono:

- **Amministratore di Hadoop:** esegue tutte le operazioni, tranne la creazione e l'eliminazione di bucket
- **Power User di Hadoop:** esegue tutte le operazioni, tranne la creazione e l'eliminazione di bucket e l'eliminazione di object
- **Utente read-only di Hadoop:** può essere visualizzare e leggere object

Per ulteriori informazioni su ECS IAM, vedere ECS IAM a pagina 37.

3.3.2.2 Supporto di client ECS HDFS

ECS è stato integrato con Ambari, che consente di implementare facilmente il file jar del client ECS HDFS e specificare ECS HDFS come file system predefinito in un cluster Hadoop. Il file jar viene installato in ogni nodo all'interno di un cluster Hadoop partecipante. ECS fornisce funzionalità di file system e storage equivalenti a quelle eseguite dai nomi e dai nodi di dati in un deployment Hadoop. ECS semplifica il flusso di lavoro di Hadoop eliminando la necessità di migrazione dei dati in un DAS Hadoop locale e/o creando un minimo di tre copie. La Figura 5 riportata di seguito mostra il file jar del client ECS HDFS installato su ogni nodo di elaborazione Hadoop e il flusso generale delle comunicazioni.

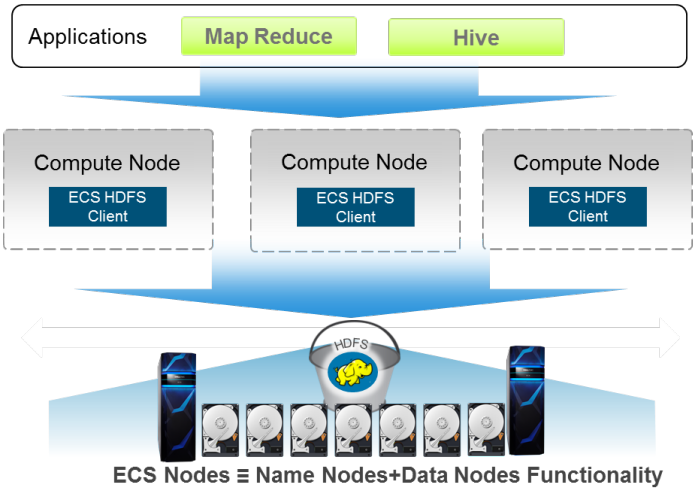

Figura 5 ECS che funge da nome e nodi di dati per un cluster Hadoop

Altri miglioramenti aggiunti in ECS per HDFS sono i seguenti:

- **Autenticazione utente proxy:** furto di identità per Hive, HBase e Oozie.
- **Sicurezza:** applicazione di ACL lato server e aggiunta di un super-user e un gruppo di super-user Hadoop nonché del gruppo predefinito nei bucket.

3.3.3 NFS

ECS include il supporto nativo dei file con NFSv3. Le principali funzioni del data service file dei file NFSv3 includono:

- **Namespace globale:** accesso ai file da qualsiasi nodo in qualsiasi sito.
- **Blocco globale:** in NFSv3 il blocco è **solo consultivo**. ECS supporta implementazioni client conformi che consentono blocchi condivisi ed esclusivi, basati su intervalli e obbligatori.
- **Accesso multiprotocollo:** accesso ai dati utilizzando diversi metodi di protocollo.

Le esportazioni, le autorizzazioni e i group mapping degli utenti di NFS vengono creati utilizzando la WebUI o l'API. I client conformi a NFSv3 eseguono il mounting delle esportazioni utilizzando i nomi di namespace e bucket. Di seguito è riportato un comando di esempio per eseguire il mounting di un bucket:

```
mount -t nfs -o vers=3 s3.dell.com:/namespace/bucket
```

Per ottenere la trasparenza del client durante un guasto del nodo, è consigliabile un servizio di bilanciamento del carico per questo flusso di lavoro.

ECS ha strettamente integrato le altre implementazioni di server, tra cui *lockmgr*, *statd*, *nfsd* e *mountd*, pertanto questi servizi non dipendono dal livello dell'infrastruttura (sistema operativo host) da gestire. Il supporto di NFSv3 presenta le seguenti caratteristiche:

- Nessun limite di progettazione per il numero di file o directory.
- La dimensione di scrittura del file può essere fino a 16 TB.
- Possibilità di scalare su un massimo di 8 siti con un singolo namespace o esportazione.
- Supporto dell'autenticazione Kerberos e AUTH_SYS.

I servizi file NFS elaborano le richieste NFS provenienti dai client; tuttavia, i dati vengono archiviati come object all'interno di ECS. Un handle di file NFS viene mappato a un ID di object. Poiché il file è fondamentalmente mappato a un object, NFS dispone di funzioni come data service di object, tra cui:

- Gestione delle quote a livello di bucket.
- Crittografia a livello di object.
- Write-Once-Read-Many (WORM) a livello di bucket.
  - WORM viene implementato utilizzando il periodo di commit automatico durante la creazione del nuovo bucket.
  - WORM è applicabile solo ai bucket non conformi.

3.3.4 Connettori e gateway

Diversi prodotti software di terze parti sono in grado di accedere all'object storage ECS. Fornitori di software indipendenti (ISV), quali Panzura, Ctera e Syncplicity, creano un livello di servizi che offrono accesso client all'object storage di ECS tramite protocolli tradizionali, come SMB/CIFS, NFS e iSCSI. Le organizzazioni possono inoltre accedere o caricare i dati nello storage ECS con i seguenti prodotti Dell EMC:

- **Isilon CloudPools:** tiering dei dati basato su policy in ECS da Isilon.
- **Data Domain Cloud Tier:** tiering nativo automatizzato dei dati deduplicati in ECS da Data Domain per la retention a lungo termine. Data Domain Cloud Tier offre una soluzione sicura e a costi contenuti per crittografare i dati nel cloud con un ingombro di storage e larghezza di banda di rete ridotti.
- **GeoDrive:** servizio di storage ECS basato su stub per desktop e server Microsoft® Windows®.

3.4 Engine di storage

Al core di ECS si trova l'engine di storage. Il livello dell'engine di storage contiene i componenti principali responsabili dell'elaborazione delle richieste, nonché dell'archiviazione, del recupero, della protezione e della replica dei dati.

In questa sezione vengono descritti i principi di progettazione e il modo in cui i dati vengono rappresentati e gestiti internamente.

3.4.1 Servizi di storage

L'engine di storage ECS include i seguenti servizi, come illustrato nella Figura 6.

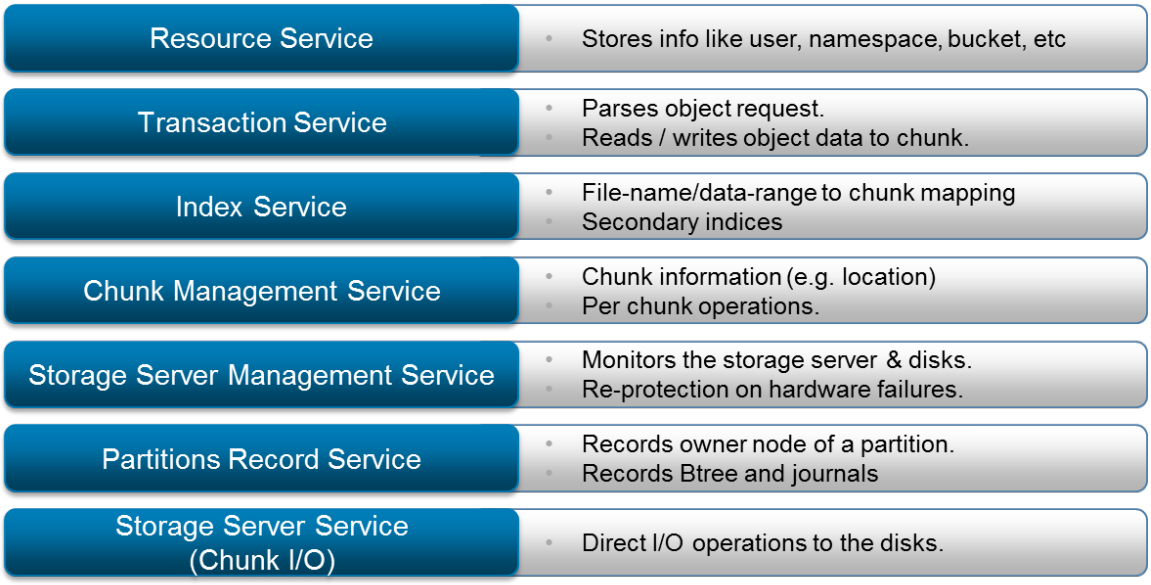

Figura 6 Servizi dell'engine di storage

I servizi dell'engine di storage sono incapsulati all'interno di un container Docker che viene eseguito su ogni nodo di ECS per fornire un servizio distribuito e condiviso.

3.4.2 Dati

I tipi primari di dati archiviati in ECS possono essere riepilogati come segue:

- **Dati:** contenuto archiviato a livello di applicazione o utente, ad esempio un'immagine. Il termine "dati" viene utilizzato come sinonimo di object, file o contenuto. Le applicazioni possono archiviare una quantità illimitata di metadati personalizzati con ogni object. L'engine di storage scrive insieme i dati e i metadati personalizzati forniti dall'applicazione associata in un repository logico. I metadati personalizzati sono una funzione avanzata dei moderni sistemi di storage che forniscono ulteriori informazioni o classificazione dei dati archiviati. I metadati personalizzati vengono formattati come coppie chiave-valore e forniti con richieste di scrittura.
- **Metadati di sistema:** informazioni e attributi di sistema relativi ai dati dell'utente e alle risorse di sistema. I metadati di sistema possono essere ampiamente classificati come segue:
  - **Identificatori e descrittori:** un insieme di attributi utilizzati internamente per identificare gli object e le relative versioni. Gli identificatori sono ID numerici o valori hash che non vengono utilizzati al di fuori del contesto del software ECS. I descrittori definiscono informazioni quali il tipo di codifica.

- **Chiavi di crittografia in formato crittografato:** le chiavi di crittografia dei dati sono considerate metadati di sistema. Sono archiviate in forma crittografata all'interno della struttura di tabelle di directory core.
- **Flag interni:** un insieme di indicatori utilizzati per tenere traccia dell'abilitazione di aggiornamenti o crittografia dell'intervallo di byte nonché per coordinare la memorizzazione nella cache e l'eliminazione.
- **Informazioni sulla posizione:** insieme di attributi con indice e posizione dei dati, ad esempio offset di byte.
- **Timestamp:** insieme di attributi che tiene traccia del tempo, ad esempio per la creazione o l'aggiornamento di un object.
- **Informazioni di configurazione/tenancy:** controllo degli accessi a namespace e object.

I dati e i metadati di sistema vengono scritti in *blocchi* su ECS. Un blocco ECS è un container logico di 128 MB di spazio contiguo. Ogni blocco può avere dati da object diversi, come illustrato di seguito nella Figura 7. ECS utilizza l'indicizzazione per tenere traccia di tutte le parti di un object che possono essere distribuite su diversi blocchi e nodi.

I blocchi vengono scritti secondo un modello "append-only" o di aggiunta in coda Il comportamento append-only significa che la richiesta di un'applicazione di modificare o aggiornare un object esistente non modificherà o eliminerà i dati scritti in precedenza in un blocco, ma piuttosto le nuove modifiche o aggiornamenti verranno scritti in un nuovo blocco. Pertanto, non è richiesto alcun blocco per i /O e non è necessaria alcuna invalidazione della memoria cache. La progettazione append-only semplifica anche il controllo delle versioni dei dati. Le versioni precedenti dei dati vengono mantenute in blocchi precedenti. Se il controllo delle versioni di S3 è abilitato ed è necessaria una versione precedente dei dati, è possibile recuperarli o ripristinarli a una versione precedente usando l'API REST di S3.

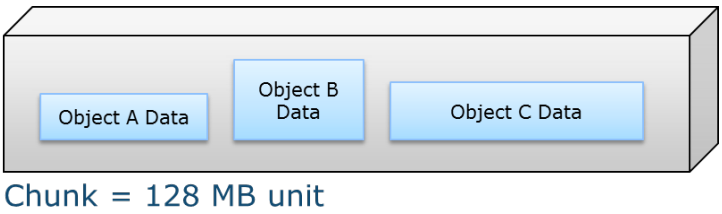

Figura 7 Blocco da 128 MB in cui sono archiviati i dati di tre object

Nella sezione *Integrità e protezione dei dati* riportata di seguito viene illustrata la modalità di protezione dei dati a livello di blocco.

3.4.3 Gestione dei dati

ECS utilizza un set di tabelle logiche per archiviare le informazioni relative agli object. Le coppie chiave-valore vengono infine archiviate su disco in una struttura B+ per un'indicizzazione rapida delle posizioni dei dati. Archiviando la coppia chiave-valore in una struttura ricercata e bilanciata, ad esempio una struttura B+, è possibile accedere rapidamente alla posizione dei dati e dei metadati. ECS implementa una struttura di unione basata su log a due livelli in cui sono presenti due strutture simili a un albero; un albero più piccolo è in memoria (tabella di memoria) e l'albero B+ principale si trova sul disco. La ricerca di coppie chiave-valore si verifica prima in memoria, successivamente nella struttura B+ principale sul disco, se necessario. Le voci in queste tabelle logiche vengono prima registrate nei log del journal e questi log vengono scritti nei dischi in blocchi con mirroring triplo. I journal vengono utilizzati per tenere traccia delle transazioni non ancora confermate nella struttura B+. Dopo aver registrato ogni transazione in un journal, la tabella in memoria viene aggiornata. Una volta che la tabella nella memoria diventa piena o dopo un certo periodo, viene unita in modo ordinato o ne viene eseguito il dump nella struttura B+ sul disco. Il numero di blocchi del journal utilizzati dal sistema è insignificante rispetto ai blocchi della struttura B+. La Figura 8 illustra questo processo.

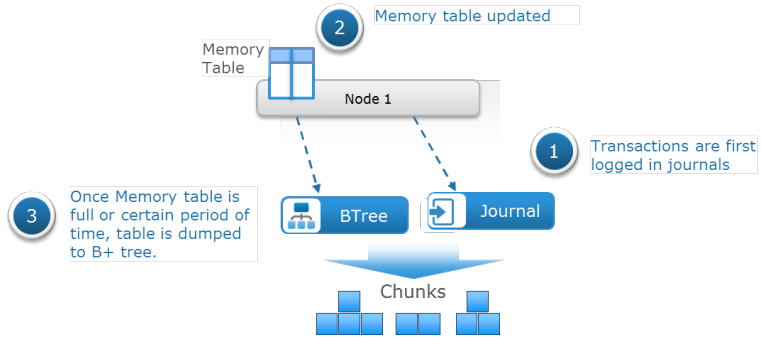

Figura 8 Tabella della memoria di cui è stato effettuato il dump nella struttura B+

Le informazioni archiviate nella tabella Object (OB) sono riportate di seguito nella Tabella 2. La tabella OB contiene i nomi degli object e la posizione del loro blocco in un determinato offset e lunghezza all'interno di tale blocco. In questa tabella, il nome dell'object è la chiave per l'indice e il valore è la posizione del blocco. Il livello di indice all'interno dell'engine di storage è responsabile del mapping da nome a blocco dell'object.

Tabella 2 Voci della tabella Object

Nome object	Posizione del blocco
ImgA	• C1:offset:length
FileB	• C2:offset:length • C3:offset:length

La tabella dei blocchi (CT) riporta la posizione di ciascun blocco come descritto in dettaglio nella Tabella 3.

Tabella 3 Voci della tabella dei blocchi

ID blocco	Sede
C1	• Node1:Disk1:File1:Offset1:Length • Node2:Disk2:File1:Offset2:Length • Node3:Disk2:File6:Offset:Length

ECS è stato progettato per essere un sistema distribuito in modo tale che lo storage e l'accesso dei dati siano distribuiti su tutti i nodi. Le tabelle utilizzate per gestire i dati e i metadati degli object aumentano nel tempo man mano che lo storage viene utilizzato e cresce. Le tabelle sono suddivise in partizioni e assegnate a nodi diversi in cui ogni nodo diventa il proprietario delle partizioni che ospita per ognuna delle tabelle. Per ottenere la posizione di un blocco, ad esempio, nella tabella Partition Records (PR) viene eseguita una query per il nodo proprietario che conosce la posizione del blocco. Una tabella PR di base è illustrata nella Tabella 4 riportata di seguito.

Tabella 4 Voci della tabella Partition Records

ID partizione	Owner
P1	Nodo 1
P2	Nodo 2
P3	Nodo 3

Se un nodo si arresta, gli altri nodi assumono la proprietà delle relative partizioni. Le partizioni vengono ricreate leggendo la radice della struttura B+ e riproducendo i journal archiviati nel disco. La Figura 9 illustra il failover della proprietà della partizione.

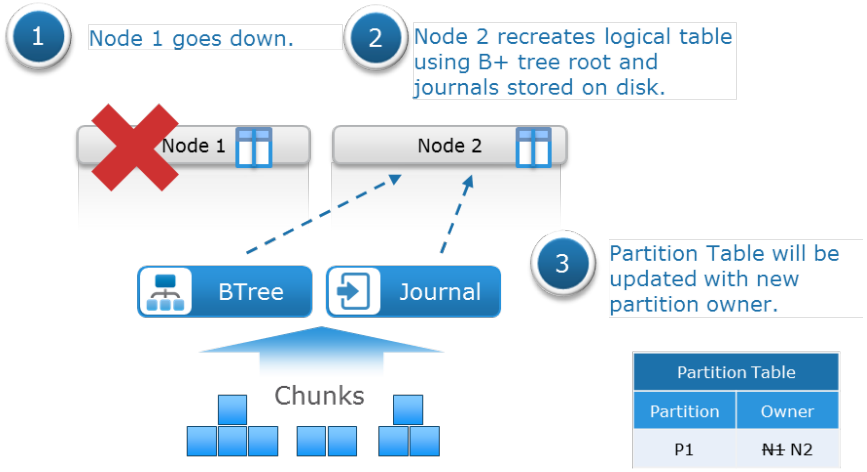

Figura 9 Failover della proprietà della partizione

3.4.4 Flusso di dati

I servizi di storage sono disponibili da qualsiasi nodo. I dati sono protetti da segmenti EC distribuiti su unità, nodi e rack. ECS esegue una funzione di checksum e archivia il risultato con ogni scrittura. Se i primi byte di dati sono comprimibili ECS comprimerà i dati. Con le letture, i dati vengono decompressi e il checksum archiviato viene convalidato. Di seguito è riportato un esempio di flusso di dati per una scrittura in cinque passaggi:

1. Il client invia la richiesta di creazione dell'object a un nodo.
2. Il nodo che serve la richiesta scrive i dati del nuovo object in un blocco di repository (abbreviazione di repository).
3. Su scrittura su disco viene eseguita correttamente, si verifica una transazione PR per immettere il nome e il percorso del blocco.
4. Il proprietario della partizione registra la transazione nei log del journal.
5. Una volta registrata la transazione nei log del journal, viene inviata una conferma al client.

La Figura 10 illustra il flusso di dati per una lettura dell'architettura di un'unità del disco rigido, come Gen2, EX300, EX500 ed EX3000:

1. Viene inviata una richiesta di lettura dell'object dal client al nodo 1.
2. Il nodo 1 utilizza una funzione hash con il nome dell'object per determinare quale nodo è il proprietario della partizione della tabella logica in cui si trovano le informazioni sull'object. In questo esempio, il nodo 2 è proprietario e pertanto il nodo 2 eseguirà una ricerca nelle tabelle logiche per ottenere la posizione del blocco. In alcuni casi, la ricerca può verificarsi in due nodi diversi, ad esempio quando il percorso non è memorizzato nella cache nelle tabelle logiche del nodo 2.
3. Dal passaggio precedente, la posizione del blocco viene fornita al nodo 1 che emetterà una richiesta di lettura di offset di byte al nodo che contiene i dati, il nodo 3 in questo esempio, e invierà i dati al nodo 1.
4. Il nodo 1 invia i dati al client richiedente.

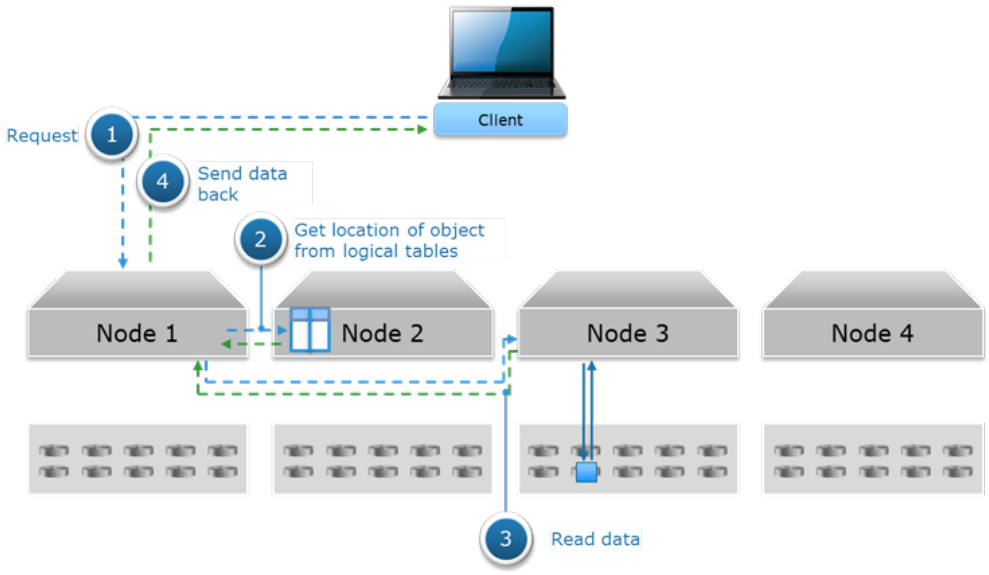

Figura 10 Lettura del flusso di dati dell'architettura di un'unità del disco rigido

La Figura 11 mostra un esempio di flusso di dati per una lettura di un'architettura All-Flash, ad esempio EXF900:

1. Viene inviata una richiesta di lettura dell'object dal client al nodo 1.
2. Il nodo 1 utilizza una funzione hash con il nome dell'object per determinare quale nodo è il proprietario della partizione della tabella logica in cui si trovano le informazioni sull'object. In questo esempio, il nodo 2 è proprietario e pertanto il nodo 2 eseguirà una ricerca nelle tabelle logiche per ottenere la posizione del blocco. In alcuni casi, la ricerca può verificarsi in due nodi diversi, ad esempio quando il percorso non è memorizzato nella cache nelle tabelle logiche del nodo 2.
3. Dal passaggio precedente, la posizione del blocco viene fornita al nodo 1 che leggerà quindi i dati direttamente dal nodo 3.
4. Il nodo 1 invia i dati al client richiedente.

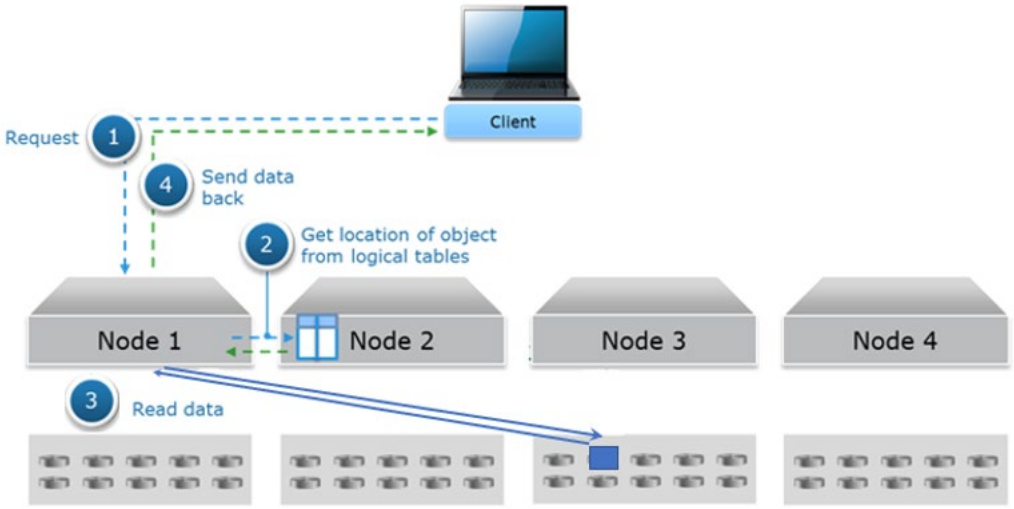

Figura 11 Lettura del flusso di dati di un'architettura All-Flash

Nota: in un'architettura All-Flash, come EXF900, ciascun nodo può leggere i dati direttamente da un altro nodo, ad eccezione dell'architettura di unità del disco rigido, nella quale ciascun nodo può leggere solo l'archivio dati contenuto in sé stesso.

3.4.5 Ottimizzazioni della scrittura per la dimensione file

Per le scritture più piccole nello storage, ECS utilizza un metodo denominato *box-carting* per ridurre al minimo l'impatto sulle prestazioni. Il box-carting aggrega una serie di scritture più piccole, di dimensioni pari a o minori di 2 MB di memoria e le scrive su disco in un'unica operazione. Il box-carting limita il numero di round trip su disco richiesti per elaborare le singole scritture.

Per le scritture di object più grandi, i nodi all'interno di ECS possono elaborare contemporaneamente le richieste di scrittura per lo stesso object e sfruttare le scritture simultanee su più spindle nel cluster ECS. In questo modo, ECS può acquisire e archiviare object di piccole e grandi dimensioni in modo efficiente.

3.4.6 Recupero dello spazio

La scrittura di blocchi in modalità append-only significa che i dati vengono aggiunti o aggiornati mantenendo innanzitutto i dati scritti originali e in secondo luogo creando nuovi segmenti di blocco netti che possono essere inclusi o meno nel container di blocchi dell'object originale. Il vantaggio della modifica dei dati append-only è un modello di accesso ai dati attivo/attivo che non è ostacolato da problemi di blocco dei file dei file system tradizionali. In questo caso, quando gli object vengono aggiornati o eliminati, i dati nei blocchi non diventano più necessari o non viene più fatto riferimento a essi. Due metodi di garbage collection utilizzati da ECS per recuperare spazio da blocchi completi eliminati o da blocchi contenenti una combinazione di frammenti di object eliminati e non eliminati a cui non viene più fatto riferimento sono i seguenti:

- **Garbage collection normale:** quando un intero blocco è garbage, recupera lo spazio.
- **Garbage collection parziale per unione:** quando un blocco è per 2/3 garbage, recupera il blocco riunendo le sue parti valide con quelle di altri blocchi parzialmente riempiti in un nuovo blocco, recuperando così dello spazio.

La garbage collection è stata applicata anche all'API di accesso ai data service ECS CAS per pulire i BLOB orfani. I BLOB orfani, che sono BLOB senza riferimenti identificati nei dati CAS archiviati in ECS, saranno idonei per il recupero dello spazio tramite i normali metodi di garbage collection.

3.4.7 Caching dei metadati SSD

I metadati ECS vengono archiviati in strutture B. Ogni struttura B può avere voci in memoria, nelle transazioni di journal e su disco. Affinché il sistema possa disporre di un quadro completo di una particolare struttura B, vengono eseguite query su tutte e tre le posizioni, che spesso includono più finestre di ricerca su disco.

Per ridurre al minimo la latenza per le ricerche di metadati, in ECS 3.5 è stato implementato un meccanismo di caching opzionale basato su SSD. La cache contiene le pagine della struttura B, alle quali sono stati effettuati accessi recenti. Ciò significa che le operazioni di lettura sulle ultime strutture B verranno effettuate sempre nella cache basata su SSD, evitando così di coinvolgere i dischi meccanici.

Di seguito sono riportati alcuni vantaggi della nuova funzionalità di memorizzazione nella cache dei metadati SSD:

- Miglioramento della latenza di lettura a livello di sistema e TPS (transazioni al secondo) per i file di piccole dimensioni
- Una unità Flash da 960 GB per nodo
- I nodi Net New in produzione includono l'unità SSD come opzione
- I nodi di campo esistenti, Gen3 e Gen2, possono essere aggiornati tramite kit di upgrade e installazione self-service
- È possibile aggiungere unità SSD mentre ECS è online
- Miglioramento dei carichi di lavoro di analisi dei file di piccole dimensioni che richiedono letture rapide di data set di grandi dimensioni
- Tutti i nodi di un VDC devono disporre di SSD per abilitare questa funzione

La fabric ECS rileva la presenza di un kit SSD installato. In questo modo il sistema viene inizializzato automaticamente e inizia a utilizzare la nuova unità. La Figura 12 mostra la memoria cache SSD abilitata.

Figura 12 Cache SSD abilitata

La memorizzazione nella cache dei metadati SSD migliora le piccole letture e la visualizzazione degli elenchi di bucket. Come testato nel nostro laboratorio, le prestazioni della visualizzazione degli elenchi migliorano del 50% con gli object da 10 MB. Le prestazioni della lettura migliorano del 35% con gli object da 10 KB e del 70% con gli object da 100 KB.

3.4.8 DVR cloud

ECS supporta la funzionalità DVR (Digital Video Recording) cloud che soddisfa i requisiti legali di copyright per le aziende fornitrici di servizi via cavo e satellite. Il requisito è che ogni unità di registrazione mappata a un object in ECS deve essere copiata per un numero predeterminato di volte. Il numero predeterminato di copie è noto come fan-out. Il numero predeterminato di copie (fan-out) non è in realtà un requisito per la ridondanza o l'aumento delle prestazioni, ma è più un requisito legale di copyright per le aziende fornitrici di servizi via cavo e satellite. ECS supporta:

- Definizione di un fan-out di copie di object create in ECS
- La lettura di una copia specifica
- L'eliminazione di una copia specifica
- L'eliminazione di tutte le copie
- La copia di una copia specifica
- L'elenco delle copie
- L'elenco di bucket degli object fan-out

La funzione DVR cloud può essere abilitata tramite la Service Console. La prima volta, è necessario abilitare la funzione DVR cloud utilizzando la Service Console. Dopo essere stato abilitato la prima volta, per impostazione predefinita DVR cloud sarà abilitato per tutti i nuovi nodi.

Eseguire il comando riportato di seguito nella Service Console per abilitare la funzione DVR cloud:

```
service-console run Enable_CloudDVR
```

La funzione DVR cloud supporta le API. Per maggiori dettagli, vedere la *Guida all'accesso ai dati ECS* per ulteriori dettagli

3.5 Fabric

Il livello Fabric fornisce clustering, integrità del sistema, gestione del software, gestione della configurazione, funzionalità di upgrade e avvisi. È responsabile dell'esecuzione di servizi e della gestione di risorse quali dischi, container e rete. Tiene traccia e reagisce alle modifiche dell'ambiente, ad esempio il rilevamento degli errori, e fornisce avvisi relativi all'integrità del sistema. Il livello Fabric presenta i seguenti componenti:

- **Agent del nodo:** gestisce le risorse host (dischi, rete, container e così via) e i processi di sistema.
- **Strumento di gestione del ciclo di vita:** gestione del ciclo di vita delle applicazioni che comporta l'avvio di servizi, ripristino, notifica e rilevamento degli errori.
- **Strumento di gestione della persistenza:** coordina e sincronizza l'ambiente distribuito ECS.
- **Registro:** archivio delle immagini di Docker per il software ECS.
- **Libreria degli eventi:** contiene l'insieme di eventi che si verificano nel sistema.
- **Strumento di gestione dell'hardware:** fornisce lo stato, le informazioni sugli eventi e il provisioning del livello hardware ai servizi di livello superiore. Questi servizi sono stati integrati per supportare l'hardware di uso comune.

3.5.1 Agent del nodo

L'agent del nodo è un agent leggero compilato in Java che viene eseguito in modo nativo su tutti i nodi ECS. I suoi compiti principali includono la gestione e il controllo delle risorse host (container, dischi, firewall, rete di Docker) e il monitoraggio dei processi del sistema. Esempi di gestione includono la formattazione e il mounting dei dischi, l'apertura delle porte richieste, la verifica dell'esecuzione di tutti i processi e la determinazione delle interfacce di rete pubbliche e private. Dispone di un flusso di eventi che fornisce eventi ordinati a un'utilità di gestione del ciclo di vita per indicare gli eventi che si verificano nel sistema. Una CLI è utile per diagnosticare i problemi ed esaminare lo stato generale del sistema.

3.5.2 Strumento di gestione del ciclo di vita

Lo strumento di gestione del ciclo di vita viene eseguito su un sottoinsieme di tre o cinque nodi e gestisce il ciclo di vita delle applicazioni in esecuzione sui nodi. Ogni strumento di gestione del ciclo di vita è responsabile del rilevamento di più nodi. Il suo obiettivo principale è gestire l'intero ciclo di vita dell'applicazione ECS dall'avvio al deployment, inclusi il rilevamento degli errori, il ripristino, la notifica e la migrazione. Esamina i flussi dell'agent del nodo e fa sì che l'agent gestisca la situazione. Quando un nodo è inattivo, risponde a guasti o incoerenze nello stato del nodo riportando il sistema a uno stato valido noto. Se un'istanza di gestione del ciclo di vita è inattiva, un'altra prende il suo posto.

3.5.3 Registro

Il registro contiene le immagini di ECS Docker utilizzate durante l'installazione, l'upgrade e la sostituzione dei nodi. Un container di Docker denominato *fabric-registry* viene eseguito su un nodo all'interno del rack ECS e contiene il repository delle immagini di ECS Docker e delle informazioni richieste per le installazioni e gli upgrade. Anche se il registro è disponibile su un nodo alla volta, tutte le immagini di Docker vengono archiviate nella cache locale in ogni nodo, pertanto una qualsiasi di esse può servire il registro.

3.5.4 Libreria degli eventi

La libreria degli eventi viene utilizzata all'interno del livello Fabric per esporre il ciclo di vita e i flussi di eventi dell'agent del nodo. Gli eventi generati dal sistema vengono mantenuti nella memoria condivisa e sul disco per fornire informazioni cronologiche sullo stato e sull'integrità del sistema ECS. Questi flussi di eventi ordinati possono essere utilizzati per ripristinare il sistema a uno stato specifico riproducendo gli eventi ordinati archiviati. Alcuni esempi di eventi includono eventi del nodo quali avviato, arrestato o danneggiato.

3.5.5 Strumento di gestione dell'hardware

Lo strumento di gestione dell'hardware è integrato nell'agent del fabric per supportare l'hardware standard del settore. Il suo scopo principale è fornire lo stato specifico dell'hardware e le informazioni sugli eventi, oltre al provisioning del livello hardware ai servizi di livello superiore in ECS.

3.6 Infrastruttura

Nei nodi dell'appliance ECS è attualmente in esecuzione SUSE Linux Enterprise Server 12 per l'infrastruttura. Per il software ECS implementato su hardware standard del settore personalizzato, il sistema operativo può anche essere RedHat Enterprise Linux o CoreOS. I deployment personalizzati vengono eseguiti tramite un formale processo di richiesta e convalida. Docker viene installato nell'infrastruttura per implementare i livelli ECS incapsulati. Il software ECS è compilato in Java, quindi Java Virtual Machine viene installato come parte dell'infrastruttura.

3.6.1 Docker

ECS viene eseguito sul sistema operativo come applicazione Java ed è incapsulato all'interno di diversi container di Docker. I container sono isolati ma condividono le risorse e l'hardware del sistema operativo sottostante. Alcune parti del software ECS vengono eseguite su tutti i nodi, mentre le altre su uno o alcuni nodi. I componenti in esecuzione in un container di Docker includono:

- **object-main:** contiene le risorse e i processi relativi ai data service, all'engine di storage nonché al portale e ai servizi di provisioning. Viene eseguito su tutti i nodi in ECS.
- **fabric-lifecycle:** contiene i processi, le informazioni e le risorse necessari per il monitoraggio a livello di sistema, la gestione della configurazione e la gestione dell'integrità. Un numero dispari di istanze fabric-lifecycle sarà sempre in esecuzione. Ad esempio, saranno in esecuzione tre istanze in un sistema a quattro nodi e cinque istanze per un sistema a otto nodi.
- **fabric-zookeeper:** servizio centralizzato per il coordinamento e la sincronizzazione di processi distribuiti, informazioni di configurazione, gruppi e servizi di denominazione. Viene definito come strumento di gestione della persistenza e viene eseguito su un numero dispari di nodi, ad esempio cinque in un sistema a otto nodi.
- **fabric-registry:** registro delle immagini di ECS Docker. Viene eseguita una sola istanza per ogni rack ECS.

Esistono altri processi e strumenti che vengono eseguiti in un container Docker, ovvero l'agent del nodo Fabric e gli strumenti del livello di astrazione hardware. La Figura 13 qui di seguito riporta un esempio di come i container ECS possono essere eseguiti in un deployment a otto nodi.

Figura 13 Esempio di container e agent di Docker in un deployment a otto nodi

La Figura 14 riporta l'output della riga di comando dopo l'immissione del comando *docker ps* su un nodo che mostra i quattro container utilizzati da ECS all'interno di Docker. Viene visualizzato un elenco con tutti i servizi correlati agli object disponibili nel sistema.

```
admin@hop-u300-11-pub-01:~> sudo docker ps
CONTAINER ID        IMAGE                                                COMMAND                  CREATED             STATUS
      PORTS               NAMES
7ba30ce42be2        ecs-monitoring/telegraf:3.5.0.0-825.b6b07cf9         "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-telegraf
e22513635cab        ecs-monitoring/grafana:3.5.0.0-825.b6b07cf9          "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-grafana
ee9db1ea40bc        emcvipr/object:3.5.0.0-120417.6a358e139f1            "/opt/vipr/boot/boot…"   5 weeks ago         Up 5 weeks
                          object-main
d11a7acd55e5        ecs-monitoring/throttler:3.5.0.0-825.b6b07cf9        "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-throttler
f94026797bb3        ecs-monitoring/fluxd:3.5.0.0-825.b6b07cf9            "/entrypoint.sh "        5 weeks ago         Up 5 weeks
                          object-fluxd
c7b8530a8bb9        caspian/fabric:3.5.0.0-4076.7d40a97                  "./boot.sh lifecycle"    5 weeks ago         Up 5 weeks
                          fabric-lifecycle
bffdf8836853        caspian/fabric-zookeeper:3.5.6.0-99.0354df7          "./boot.sh 1 1=169.2…"   5 weeks ago         Up 5 weeks
                          fabric-zookeeper
f4420f7f7d51        caspian/fabric-registry:2.3.1.0-68.10d1aca           "/opt/docker-registr…"   5 weeks ago         Up 5 weeks
                          fabric-registry
admin@hop-u300-11-pub-01:~> sudo dockobj
hop-u300-11-pub-01:/ # cd /opt/storageos/
hop-u300-11-pub-01:/opt/storageos # ls bin/*svc
bin/blobsvc     bin/coordinatorsvc  bin/eventsvc  bin/objcontrolsvc  bin/storagemanagementsvc
bin/cassvc      bin/dataheadsvc     bin/filesvc   bin/objheadsvc     bin/syssvc
bin/controlsvc  bin/ecsportalsvc    bin/hdfssvc   bin/resourcesvc    bin/transformsvc
```

Figura 14 Processi, risorse, strumenti e file binari nel container object-main

4 Modelli hardware dell'appliance

I punti di ingresso flessibili consentono a ECS di scalare rapidamente i dati nell'ordine di petabyte ed exabyte. Con un impatto aziendale minimo, una soluzione ECS può essere scalata in modo lineare sia in termini di capacità che di prestazioni aggiungendo nodi e dischi.

I modelli hardware dell'appliance ECS sono caratterizzati dalla generazione di hardware. La serie di appliance di terza generazione, nota come Gen3 o serie EX, include tre modelli hardware. In questa sezione viene fornita una panoramica generale della serie EX. Per i dettagli completi, consultare la *Guida all'hardware della ECS serie EX*.

Le informazioni sull'hardware dell'appliance ECS di prima e seconda generazione sono disponibili nella *Guida all'hardware Dell EMC ECS serie D e U*.

4.1 Serie EX

I modelli di appliance della serie EX sono basati su server e switch Dell standard. Le offerte della serie sono le seguenti:

- **EX300:** EX300 ha una capacità raw iniziale di 60 TB. È la piattaforma di storage perfetta per le app native per il cloud e le iniziative di Digital Transformation dei clienti. EX300 è ideale per la modernizzazione dei deployment Centera. Ancora più importante, EX300 può scalare a capacità più grandi a costi contenuti. Fornisce 12 unità per nodo e opzioni per dischi da 1 TB, 2 TB, 4 TB, 8 TB e 16 TB (ma tutti uguali nello stesso nodo)
- **EX500:** EX500 è l'appliance di ultima generazione che punta a fornire economia in termini di densità. Con opzioni per 12 o 24 unità, opzioni per dischi da 8 TB, 12 TB e 16 TB (ma tutti uguali nello stesso nodo). Intervallo di cluster da 480 TB a 6,1 PB per rack. Questa serie offre un'opzione versatile per le aziende di medie dimensioni che desiderano supportare i moderni casi d'uso di applicazioni e/o archivi profondi.
- **EX3000:** EX3000 ha una capacità massima di 11,5 PB di storage raw per rack, da 30 a 90 unità per nodo, dischi da 12 TB o 16 TB e può crescere in termini di exabyte in diversi siti, fornendo una soluzione di data center completa e scalabile ideale per carichi di lavoro con ingombri di dati di dimensioni maggiori. Questi nodi sono disponibili in due diverse configurazioni note come EX3000S ed EX3000D. EX3000S è un singolo nodo, mentre EX3000D è uno chassis a doppio nodo. Questi nodi ad alta densità sono dischi sostituibili a caldo. Iniziano con un minimo di trenta dischi per nodo. Trenta unità per nodo ECS è il punto in cui diminuiscono i miglioramenti in termini di prestazioni con l'aggiunta di più unità. Con trenta o più unità in ogni nodo come minimo, le aspettative di prestazioni sono simili in ogni nodo EX3000 indipendentemente dal numero di unità.
- **EXF900:** EXF900 è una soluzione di object storage All-Flash di nodi hyper-converged per deployment di ECS a bassa latenza ed elevati IOPS. Con opzioni per 12 o 24 unità, opzioni per unità SSD NVMe da 3,84 TB (il drive SSD NVMe da 7,68 TB sarà supportato non appena sarà disponibile l'hardware). Questa piattaforma parte da una configurazione RAW minima di 230 TB ed è scalabile fino a 1,4 PB RAW per rack. La Figura 15 mostra un nodo di EXF900.

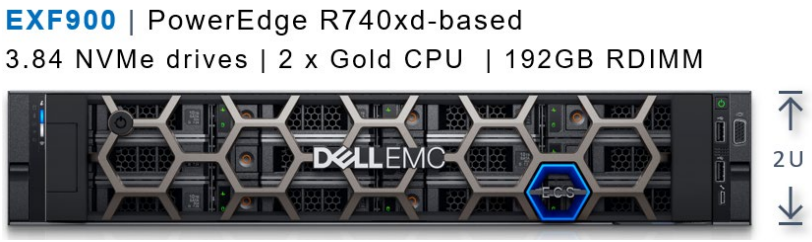

Figura 15 Nodo EXF900

Nota: la funzione SSD Read Cache non si applica a EXF900; la funzione DVR cloud non è supportata su EXF900; il Tech Refresh non è supportato con EXF900; EXF900 non può coesistere con altri hardware non EXF900 in un VDC; EXF900 non può coesistere con altri hardware non EXF900 in GEO (tutti i siti devono essere EXF900).

Le opzioni di capacità di avvio della serie EX consentono ai clienti di avviare un deployment ECS con la sola capacità necessaria e di crescere facilmente in base alle esigenze in futuro. Vedere le *Specifiche tecniche degli appliance ECS* per ulteriori dettagli sugli appliance della serie EX e anche sui precedenti appliance serie U e D Gen2.

Gli aggiornamenti post-deployment ai nodi della serie EX non sono supportati. tra cui:

- Modifica della CPU.
- Regolazione della capacità di memoria.
- Upgrade delle dimensioni del disco rigido.

4.2 Rete di appliance

A partire dal rilascio degli appliance della serie EX, viene utilizzata una coppia ridondante di switch di gestione back-end dedicati. Passando al nuovo appliance, ECS è ora in grado di adottare una modalità di switch della configurazione front-end e back-end.

Gli appliance EX300, EX500 ed EX3000 utilizzano tutti Dell EMC S5148F per la coppia di switch di front-end e back-end. L'appliance EXF900 utilizza Dell EMC S5248F per la coppia di switch di front-end e back-end e S5232F per lo switch di aggregazione di back-end. Si noti che i clienti hanno la possibilità di utilizzare i propri switch di front-end anziché gli switch Dell EMC.

4.2.1 S5148F: switch pubblici di front-end

È possibile ottenere due switch Ethernet Dell EMC S5148F 25 GbE 1U opzionali per la connessione di rete oppure il cliente può fornire la propria coppia HA da 10 o 25 GbE per la connettività front-end. Gli switch pubblici sono spesso indicati con i nomi *"hare"* e *"rabbit"* o semplicemente front-end.

Attenzione: è necessario disporre di connessioni dalla rete del cliente a entrambi gli switch di front-end (rabbit e hare) per mantenere l'architettura high availability dell'appliance ECS. Se il cliente decide di non connetter la propria come richiesto per la modalità HA (High Availability), non sussiste alcuna garanzia di high availability dei dati per l'utilizzo di questo prodotto.

Questi switch forniscono 48 porte SFP28 da 25 GbE e 6 porte QSFP28 da 100 GbE. Ulteriori dettagli di questi due tipi di porta sono riportati di seguito:

- SFP28 è una versione migliorata di SFP
  - SFP supporta fino a 16 Gb/s, SFP28 supporta fino a 28 Gb/s
  - Stesso fattore di forma
  - Compatibilità con le versioni precedenti dei moduli SFP
- QSFP28 è una versione migliorata di QSFP+
  - QSFP+ supporta fino a 4 linee da 16 Gb/s, QSFP28 supporta fino a 4 linee da 28 Gb/s
    - > Linee aggregate di QSFP+ per ottenere Ethernet da 40 Gb/s
    - > Linee aggregate di QSFP28 per ottenere Ethernet da 100 Gb/s
  - Stesso fattore di forma
  - Compatibilità con le versioni precedenti dei moduli QSFP+
  - Può essere suddiviso in 4 singole linee di SFP28

Nota: due cavi LAG da 100 GbE vengono insieme agli switch pubblici Dell EMC S5148F da 25 GbE. Le organizzazioni che forniscono i propri switch pubblici devono fornire i cavi LAG, SFP richiesti o cavi di connessione esterni.

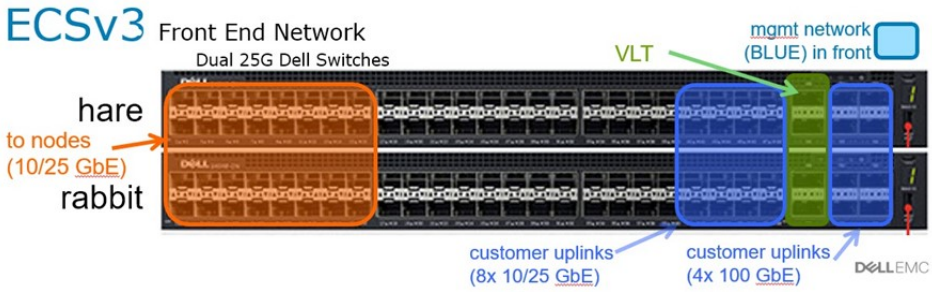

Figura 16 Designazione e uso della porta dello switch di rete di front-end

La Figura 16 precedente mostra una rappresentazione visiva del modo in cui le porte devono essere utilizzate per abilitare il traffico sui nodi ECS e sulle porte uplink dei clienti. Questo è lo standard in tutte le implementazioni.

4.2.2 S5148F: switch privati di back-end

Entrambi gli switch Ethernet Dell EMC S5148F 1U da 25 GbE richiesti con 48 porte SFP da 25 GbE e 6 porte uplink da 100 GbE sono inclusi in ogni rack ECS. Questi switch sono spesso indicati con i nomi *fox* e *hound* o switch di back-end e sono responsabili della rete di gestione. Nelle versioni ECS future, gli switch back-end forniranno anche la separazione di rete per il traffico di replica. Lo scopo principale della rete privata è la gestione remota e l'avvio PXE della console per la gestione delle installazioni e l'abilitazione della gestione e del provisioning a livello di rack e cluster. La Figura 17 mostra la vista anteriore di due switch Dell da 25 GbE.

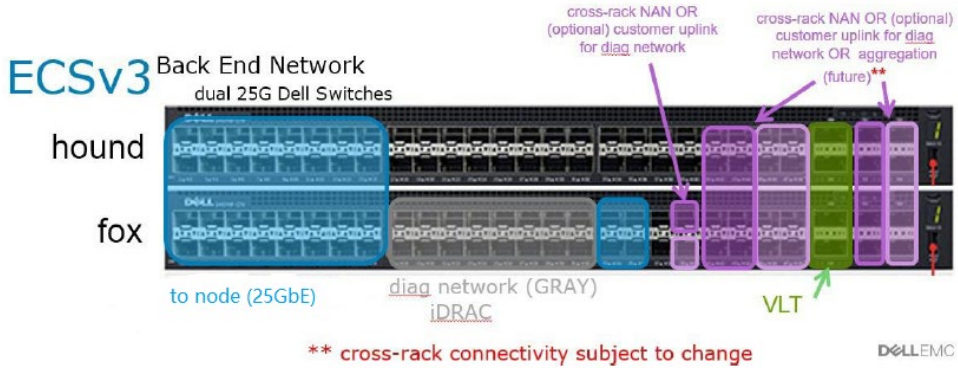

Figura 17 Designazione e uso della porta dello switch di rete di back-end

Il diagramma precedente fornisce una rappresentazione visiva del modo in cui le porte devono essere utilizzate per abilitare il traffico di gestione e le porte di diagnostica di ECS. Queste allocazioni di porte sono standard in tutte le implementazioni. Le possibili porte che verranno utilizzate in futuro sono indicate in viola; tuttavia, tale utilizzo è soggetto a modifiche in futuro.

4.2.3 S5248F: switch pubblici di front-end

Dell EMC offre una coppia HA opzionale di switch di front-end S5248F da 25 GbE per la connessione di rete del cliente al rack. Sono disponibili due cavi VLT (Virtual Link Trunking) da 200 GbE (QSFP28-DD) per coppia HA. Questi switch sono chiamati switch Hare e Rabbit. La Figura 18 fornisce una rappresentazione visiva del modo in cui le porte devono essere utilizzate per abilitare il traffico dei nodi ECS e le porte uplink dei clienti.

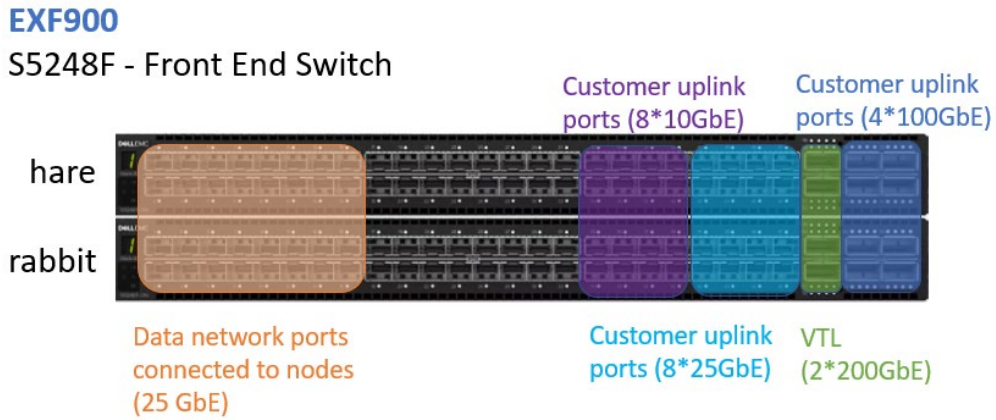

Figura 18 Designazione e uso della porta dello switch di rete di front-end

4.2.4 S5248F: switch privati di back-end

Dell EMC fornisce due switch di back-end S5248F da 25 GbE con due cavi VLT da 200 GbE (QSFP28-DD). Questi switch sono definiti switch Hound e Fox. Tutti i cavi iDRAC connessi ai nodi e tutte le connessioni dei cavi di gestione degli switch di front-end arrivano allo switch Fox. La Figura 19 fornisce una rappresentazione visiva del modo in cui le porte devono essere utilizzate per abilitare il traffico di gestione e le porte di diagnostica di ECS. Queste allocazioni di porte sono standard in tutte le implementazioni.

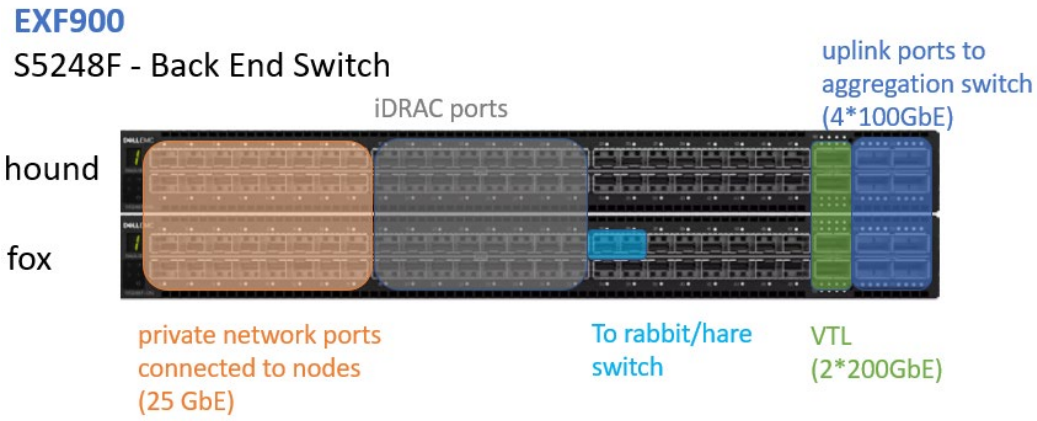

Figura 19 Designazione e uso della porta dello switch di rete di back-end

4.2.5 S5232: switch di aggregazione dei link

Dell EMC fornisce due switch di aggregazione di back-end S5232F da 100 GbE (AGG1 e AGG2) con quattro cavi VLT da 100 GbE. Questi switch sono definiti switch Falcon e Eagle. Nella Figura 20 che segue, tutte le porte contrassegnate indicano le designazioni delle porte. Questa configurazione consente di connettersi a 7 rack di nodi EXF900.

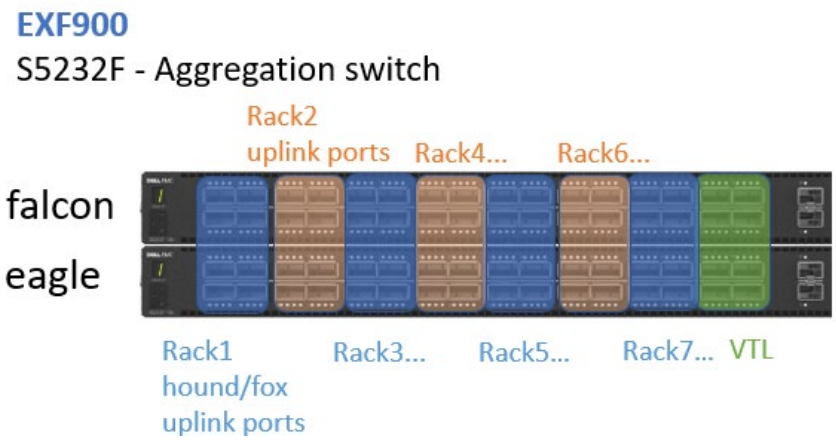

Figura 20 Designazione e utilizzo delle porte dello switch di aggregazione

Per ulteriori informazioni sulla rete e sul cablaggio, vedere la *Guida all'hardware ECS Serie EX.*

5 Separazione della rete

ECS supporta la separazione di diversi tipi di traffico di rete per la sicurezza e l'isolamento delle prestazioni. I tipi di traffico che possono essere separati includono:

- Gestione
- Replica
- Dati

Esiste una modalità di funzionamento denominata *modalità di separazione di rete*. In questa modalità ogni nodo può essere configurato a livello di sistema operativo con un massimo di tre indirizzi IP, o reti logiche, per ognuno dei diversi tipi di traffico. Questa funzione è stata progettata per fornire la flessibilità di creare tre reti logiche separate per la gestione, la replica e i dati o combinarle per creare due reti logiche; ad esempio, la gestione delle istanze e il traffico di replica rappresentano un'unica rete logica e il traffico di dati in un'altra rete logica. È possibile configurare una seconda rete di dati logica per il traffico solo CAS, consentendo la separazione del traffico CAS da altri tipi di traffico dati come S3.

L'implementazione ECS della separazione di rete richiede che ogni traffico di rete logico sia associato a servizi e porte. Ad esempio, i servizi del portale ECS comunicano tramite le porte 80 o 443, pertanto tali porte e servizi saranno collegati alla rete logica di gestione. È possibile configurare una seconda rete di dati; tuttavia, è solo per il traffico CAS. La Tabella 5 qui di seguito evidenzia i servizi associati a un tipo di rete logica. Per un elenco completo dei servizi associati alle porte, consultare la più recente *Guida alla configurazione della sicurezza di ECS*.

Tabella 5 Mapping dei servizi a una rete logica

Servizi	Rete logica	ID
WebUI e API, SSH, DNS, NTP, AD, SMTP	Gestione	public.mgmt
Dati del client	Dati	public.data
	Dati solo CAS	public.data2
Replica dei dati	Replica	public.repl
SRS (Dell EMC Secure Remote Services)	A condizione che il gateway SRS di rete sia collegato	public.data o public.mgmt

Nota: ECS 3.6 consente l'accesso ai dati S3 sulla rete dati (predefinita) e dati2 (anche se S3 non è abilitato per impostazione predefinita su dati2). Per abilitare l'accesso ai dati S3 sulla rete dati2, è richiesto il file public.data; contattare il supporto remoto ECS.

La separazione di rete è ottenibile logicamente utilizzando indirizzi IP diversi, praticamente utilizzando VLAN diverse o fisicamente utilizzando cavi diversi. Il comando *setrackinfo* viene utilizzato per configurare gli indirizzi IP e le VLAN. La configurazione VLAN a livello di switch o lato client deve essere effettuata dal cliente. Per la separazione della rete fisica, i clienti devono inviare una richiesta di qualificazione del prodotto (RPQ) contattando Dell EMC Global Business Service. Per ulteriori informazioni sulla separazione della rete, vedere il white paper *ECS Networking and Best Practices* che fornisce una vista generale della separazione della rete.

6 Sicurezza

La sicurezza ECS è implementata a livello di amministrazione, trasporto e dati. L'autenticazione dell'utente e dell'amministratore viene ottenuta tramite Active Directory, metodi LDAP, Keystone o direttamente all'interno del portale ECS. La sicurezza a livello di dati viene eseguita tramite HTTPS per i dati in movimento e/o la crittografia lato server per i dati at-rest.

6.1 Autenticazione

ECS supporta i metodi di autenticazione Active Directory, LDAP, Keystone e IAM per fornire l'accesso alla gestione e configurazione di ECS; esistono tuttavia delle limitazioni, come illustrato nella Tabella 6. Per ulteriori informazioni sulla sicurezza, consultare la *Guida alla configurazione della sicurezza di ECS* più recente.

Tabella 6 Metodi di autenticazione supportati

Metodo di autenticazione	**Supportato**
Active Directory	• Supporto del gruppo di Active Directory per gli utenti di gestione • Supporto dei gruppi di Active Directory per i metodi di self-provisioning degli utenti di object con chiavi self-service tramite API • Supporto di più domini
LDAP	• Gli utenti di gestione possono eseguire l'autenticazione singolarmente tramite LDAP • I gruppi LDAP NON sono supportati per gli utenti di gestione • LDAP è supportato per gli utenti di object (chiavi self-service tramite API) • Supporto di più domini.
Keystone	• Policy RBAC non ancora supportati. • Nessun supporto per i token senza ambito • Nessun supporto di più server Keystone per sistema ECS
IAM	• Fornisce la federazione delle identità e il Single Sign-On (SSO) tramite gli standard SAML 2.0 • Disponibile solo tramite il protocollo S3

6.2 Autenticazione dei data service

L'accesso agli object tramite le API RESTful viene protetto tramite HTTPS (TLS v1.2). Le richieste in ingresso vengono autenticate utilizzando metodi definiti, ad esempio HBAC (Hash-Based Message Authentication Code), Kerberos o l'autenticazione con token. La Tabella 7 che segue riporta i diversi metodi utilizzati per ciascun protocollo.

Tabella 7 Autenticazione dei data service

Protocolli		**Metodi di autenticazione**
Object	S3	V2 (HMAC-SHA1), V4 (HMAC-SHA256)
	Swift	Token - Keystone v2 e v3 (con ambito, UUID, token PKI), SWAuth v1
	Atmos	HMAC-SHA1
	CAS	File PEA chiave segreta
File	HDFS	Kerberos
	NFS	Kerberos, AUTH_SYS

6.3 Crittografia dei dati inattivi (D@RE)

I requisiti di conformità spesso impongono l'uso della crittografia per proteggere i dati scritti su dischi. In ECS la crittografia può essere abilitata a livello di namespace e bucket. Le funzioni chiave di ECS D@RE includono:

- Crittografia low touch nativa at-rest: facile da abilitare, configurazione semplice
- CIPHER (AES-256 CTR) utilizzati
- Crittografia con chiave pubblica RSA con una lunghezza di 2048 bit
- Supporto a livello di cluster EKM (External Key Management):
  - Gemalto SafeNet
  - IBM Security Key Lifecycle Manager
- Rotazione delle chiavi
- La semantica di crittografia S3 supporta l'utilizzo di intestazioni HTTP, ad esempio *x-amz-server-side-encryption*
- Conformità FIPS 140-2 agli standard di sicurezza della crittografia del governo degli Stati Uniti

Nota: la modalità FIPS 140-2 obbliga all'utilizzo esclusivo di algoritmi approvati in D@RE; la conformità FIPS 140-2 è valida solo per il modulo D@RE, non per l'intero prodotto ECS.

ECS utilizza una gerarchia di chiavi per crittografare e decrittografare i dati. Il gestore chiavi nativo archivia una chiave privata comune a tutti i nodi per decrittografare la chiave principale. Con la configurazione EKM, la chiave principale è fornita da EKM. EKM ha fornito le chiavi che risiedono in memoria solo su ECS. Non vengono mai archiviate nello storage persistente all'interno di ECS.

In un ambiente con replica geografica, quando un nuovo sistema ECS si unisce a una federazione esistente, la chiave principale viene estratta utilizzando la chiave pubblica-privata del sistema esistente e crittografata utilizzando la nuova coppia di chiavi pubblica-privata generata dal nuovo sistema che è stata unita alla federazione. Da questo punto in poi, la chiave principale è globale e nota a entrambi i sistemi all'interno della federazione. Quando si utilizza EKM, tutti i sistemi federati recuperano la chiave principale dal sistema di gestione delle chiavi.

6.3.1 Rotazione delle chiavi

ECS supporta la modifica delle chiavi di crittografia. Questa operazione può essere eseguita periodicamente per limitare la quantità di dati protetti da un set specifico di chiavi di KEK (Key Encryption Key) o in risposta a una potenziale perdita o compromissione. Un record KEK di rotazione viene utilizzato in combinazione con altre chiavi principali per creare chiavi di wrapping virtuali per la protezione delle DEK (Data Encryption Key) e dei KEK del namespace.

Le chiavi di rotazione vengono generate o fornite in modo nativo e gestite da un EKM. ECS utilizza la chiave di rotazione corrente per creare chiavi di wrapping virtuali al fine di proteggere qualsiasi DEK o KEK, indipendentemente dal fatto che la gestione delle chiavi venga eseguita in modo nativo o esterno.

Durante le scritture, ECS esegue il wrapping della DEK generata in modo casuale utilizzando una chiave di wrapping virtuale creata utilizzando il bucket e la chiave di rotazione attiva.

Come parte della rotazione delle chiavi, ECS esegue nuovamente il wrapping di tutti i record KEK del namespace con una nuova chiave KEK principale virtuale creata dalla nuova chiave di rotazione, il contesto segreto associato e la chiave principale attiva. Questa operazione viene eseguita per proteggere l'accesso ai dati protetti dalle chiavi di rotazione precedenti.

L'utilizzo di un EKM influisce sul percorso di lettura/scrittura degli object crittografati. La rotazione delle chiavi consente una maggiore protezione dei dati usando le chiavi di wrapping virtuali le DEK e le KEK del namespace. Le chiavi di wrapping virtuale non sono persistenti e derivano da due gerarchie indipendenti di chiavi persistenti. Con l'uso di EKM, la chiave di rotazione non viene archiviata in ECS e offre ulteriore sicurezza dei dati. Aggiungiamo principalmente nuovi record KEK e aggiorniamo gli ID attivi, ma non eliminiamo mai nulla.

Ulteriori punti da considerare per quanto riguarda la rotazione delle chiavi su ECS sono i seguenti:

- Il processo di rotazione delle chiavi modifica solo la chiave di rotazione corrente. Le chiavi principali, di namespace e bucket esistenti non cambiano durante il processo di rotazione delle chiavi.
- La rotazione delle chiavi a livello di namespace o bucket non è supportata, tuttavia l'ambito di rotazione è a livello di cluster, pertanto tutti i nuovi object crittografati di sistema saranno interessati.
- I dati esistenti non vengono nuovamente crittografati a causa delle chiavi in rotazione.
- ECS non supporta la rotazione delle chiavi in caso di interruzione dell'alimentazione.
  - Interruzione temporanea dell'alimentazione del sito (TSO) durante la rotazione: l'attività di rotazione delle chiavi viene sospesa fino al ripristino dell'alimentazione.
  - Il PSO è in corso. ECS deve uscire da un PSO prima che la rotazione della chiave venga abilitata. Se si verifica un PSO durante la rotazione, la rotazione avrà immediato esito negativo.
- La crittografia del bucket non è richiesta per eseguire la crittografia degli object tramite S3.
- I metadati degli object dei client indicizzati utilizzati come chiave di ricerca non vengono crittografati.

Per ulteriori informazioni su D@RE, EKM e rotazione delle chiavi, vedere la *Guida alla configurazione della sicurezza di ECS* più recente.

6.4 ECS IAM

ECS Identify and Access Management (IAM) consente di controllare e proteggere l'accesso alle risorse ECS S3. Questa funzionalità garantisce che ogni richiesta di accesso a una risorsa ECS sia identificata, autenticata e autorizzata. ECS IAM consente all'amministratore di aggiungere utenti, ruoli e gruppi. L'amministratore può inoltre limitare l'accesso aggiungendo policy alle entità ECS IAM.

Nota: ECS IAM può essere usata solo con S3. Non è abilitata per bucket abilitati per CAS o file system.

ECS IAM è costituita dai seguenti componenti:

- **Gestione degli account:** consente di gestire le identità IAM all'interno di ciascun namespace, ad esempio utenti, gruppi e ruoli
- **Gestione degli accessi:** l'accesso viene gestito creando policy e collegandole alle identità o risorse IAM
- **Identity Federation:** l'identità viene stabilita e autenticata tramite SAML (Security Assertion Markup Language). Una volta stabilita l'identità, utilizzare il Secure Token Service per ottenere le credenziali temporanee che verranno utilizzate per accedere alla risorsa
- **Secure Token Service:** consente di richiedere credenziali temporanee per l'accesso con più account alle risorse e anche per utenti autenticati tramite l'autenticazione SAML da un fornitore di identità aziendali o un servizio di directory

Utilizzando IAM, è possibile controllare chi è autenticato e autorizzato a utilizzare le risorse ECS creando e gestendo

- **Utenti:** l'utente IAM rappresenta una persona o un'applicazione nel namespace che può interagire con le risorse ECS
- **Gruppi:** il gruppo IAM è una raccolta di utenti IAM. Utilizzare i gruppi per specificare le autorizzazioni per un insieme di utenti IAM
- **Ruoli:** il ruolo IAM è un'identità che può essere assunta da chiunque richieda quel ruolo. Un ruolo è simile a un utente, vale a dire un'identità con policy di autorizzazione che stabiliscono cosa quell'identità può e non può fare
- **Policy:** la policy IAM è un documento in formato JSON che definisce le autorizzazioni per un ruolo. Assegnare e associare le policy a utenti IAM, gruppi IAM e ruoli IAM.
- **Fornitore di SAML:** SAML è uno standard aperto per lo scambio dei dati di autenticazione e autorizzazione tra un fornitore di identità e un fornitore di servizi. Il fornitore SAML in ECS viene utilizzato per stabilire un rapporto di fiducia tra un fornitore di identità (IdP) compatibile con SAML ed ECS

A ciascun sistema ECS viene assegnato un account ECS IAM. Questo account supporta più namespace e dispone di entità IAM correlate definite nel relativo namespace.

- I singoli namespace supportano la gestione degli account utilizzando entità IAM ECS, quali utenti, ruoli e gruppi.
- Le policy, le autorizzazioni, le ACL associate alle entità ECS IAM e le risorse ECS S3 supportano la gestione dell'accesso alle funzionalità di ECS IAM.
- ECS IAM supporta l'accesso a più account utilizzando il Security Assertion Markup Language (SAML) e i ruoli.
- ECS IAM supporta la chiave di accesso Amazon Web Services (AWS) per accedere a IAM e S3 in ECS.

Per ulteriori informazioni su ECS IAM, vedere la *Guida alla sicurezza di ECS* più recente

6.5 Etichettatura degli object

L'etichettatura degli object consente la classificazione degli object tramite l'assegnazione di tag ai singoli object. A un singolo object possono essere assegnati più tag, consentendo una classificazione multidimensionale.

Un tag potrebbe includere informazioni sensibili, come una cartella clinica, oppure potrebbe etichettare un object con un determinato prodotto classificato come riservato. L'etichettatura è una sotto-risorsa di un object che ha un ciclo di vita integrato con le operazioni di quell'object. È possibile aggiungere tag a nuovi object quando vengono caricati oppure aggiungere tag a object esistenti. È accettabile utilizzare tag per etichettare object contenenti dati riservati, quali le informazioni di identificazione personale (PII) o le informazioni sanitarie protette (PHI). I tag non devono contenere informazioni riservate, poiché possono essere visualizzati senza disporre dell'autorizzazione di lettura effettiva per un object.

6.5.1 Informazioni aggiuntive sull'etichettatura degli object

Questa sezione fornisce informazioni sull'etichettatura degli object in IAM, sull'etichettatura degli object con le policy di bucket, sulla gestione dei tag degli object durante un TSO/PSO e sull'etichettatura degli object durante la gestione del loro ciclo di vita. Di seguito sono riportate altre considerazioni:

- Etichettatura degli object in IAM
  - La funzione chiave dell'etichettatura degli object come sistema di classificazione si realizza quando è integrata con le policy IAM. Ciò consente all'amministratore di configurare autorizzazioni utente specifiche. Ad esempio, l'amministratore può aggiungere una policy che consente a tutti gli utenti di accedere agli object con un tag specificato oppure configurare e concedere autorizzazioni agli utenti che consentono loro di gestire i tag su object specifici. L'altro aspetto chiave dell'etichettatura degli object è il modo e il luogo in cui i tag vengono mantenuti. Questo è importante perché ha un impatto diretto su vari aspetti del sistema.
- Etichettatura degli object con le policy bucket
  - L'etichettatura degli object consente di classificare gli object; inoltre, l'etichettatura viene integrata con varie policy. La policy di gestione del ciclo di vita consente di effettuare la configurazione a livello di bucket. Le versioni precedenti di ECS supportano la scadenza, l'interruzione degli upload incompleti e l'eliminazione dell'indicatore di eliminazione dei tag degli object scaduti. Il filtro può includere più condizioni, tra cui una condizione basata sui tag. Ogni tag nella condizione del filtro deve corrispondere alla chiave e al valore.
- Etichettatura degli object durante un TSO/PSO
  - L'etichettatura degli object è un'altra voce impostata nei metadati di sistema, non è richiesta alcuna gestione particolare durante un TSO/PSO. Esiste un limite impostato per il numero di tag che possono essere associati a ciascun object, la dimensione dei metadati di sistema insieme all'etichettatura degli object è ben all'interno dei limiti di memoria.
- Etichettatura degli object durante la gestione del ciclo di vita degli object
  - L'etichettatura degli object fa parte dei metadati di sistema e viene gestita contemporaneamente ai metadati di sistema, durante la gestione del ciclo di vita. Lo scanner per la logica di scadenza e l'eliminazione del ciclo di vita richiede la comprensione delle policy basate sui tag. I tag degli object consentono la gestione granulare del ciclo di vita degli object, in cui è possibile specificare un filtro basato sui tag, oltre a un prefisso del nome chiave, in una regola del ciclo di vita.

Per ulteriori informazioni sull'etichettatura degli object in ECS, vedere la *Guida alla configurazione della sicurezza di ECS* più recente.

7 Integrità e protezione dei dati

Per l'integrità dei dati, ECS utilizza checksum. I checksum vengono creati durante le operazioni di scrittura e vengono archiviati con i dati. Nelle letture i checksum vengono calcolati e confrontati con la versione archiviata. Un'attività in background analizza le informazioni di checksum in modo proattivo.

Per la protezione dei dati, ECS utilizza il triplo mirroring per i blocchi del journal e schemi EC separati per i blocchi di *repo* (dati del repository utente) e *btree* (struttura B+).

La codifica di erasure offre maggiore protezione dei dati da un guasto del disco, del nodo e del rack in modo efficiente in termini di storage rispetto ai sistemi di protezione convenzionali. L'engine di storage di ECS implementa la correzione degli errori Reed Solomon utilizzando due schemi:

- 12+4 (predefinito): il blocco è suddiviso in 12 segmenti di dati. Vengono creati 4 segmenti di codifica (parità).
- 10+2 (archiviazione a freddo): il blocco viene suddiviso in 10 segmenti di dati. Vengono creati 2 segmenti di codifica.

Utilizzando il valore predefinito di 12+4, i 16 segmenti risultanti vengono distribuiti tra i nodi del sito locale. I segmenti di dati e codifica di ogni blocco sono distribuiti equamente tra i nodi del cluster. Ad esempio, con 8 nodi, ogni nodo ha 2 segmenti (su 16 totali). L'engine di storage può ricostruire un blocco da 12 dei 16 segmenti.

ECS richiede un minimo di sei nodi per l'opzione di archiviazione a freddo, in cui viene utilizzato uno schema 10+2 invece di 12+4. EC si arresta quando il numero di nodi scende al di sotto del minimo richiesto per lo schema EC.

Quando un blocco è pieno o dopo un determinato periodo, viene sigillato, viene calcolata la parità e i segmenti di codifica vengono scritti su dischi nel dominio dei guasti. I dati di blocco rimangono come una singola copia costituita da 16 segmenti (12 di dati, 4 di codice) distribuiti in tutto il cluster. ECS utilizza i segmenti di codice per la ricostruzione dei blocchi solo quando si verifica un errore.

Quando l'infrastruttura sottostante di un VDC cambia a livello di nodo o rack, i livelli Fabric rilevano la modifica e attivano uno scanner di ribilanciamento come attività in background. Lo scanner calcola il layout migliore per i segmenti EC tra domini dei guasti per ogni blocco utilizzando la nuova topologia. Se il nuovo layout offre una protezione migliore rispetto al layout esistente, ECS ridistribuisce i segmenti EC in un'attività in background. Questa attività ha un impatto minimo sulle prestazioni del sistema; tuttavia, si verificherà un aumento del traffico tra nodi durante il ribilanciamento. Si verifica anche il bilanciamento delle partizioni della tabella logica nei nuovi nodi e, in futuro, i blocchi del journal e della struttura B+ appena creati vengono allocati in modo uniforme nei nodi esistenti e in quelli nuovi. La ridistribuzione migliora la protezione locale sfruttando tutte le risorse all'interno dell'infrastruttura.

Nota: si consiglia di non attendere che la piattaforma di storage sia completamente piena prima di aggiungere ulteriori unità o nodi. Una soglia di utilizzo dello storage ragionevole è del 70% tenendo conto del tasso di acquisizione giornaliero e il tempo previsto di ordinazione, consegna e integrazione di unità/nodi aggiunti.

7.1 Conformità

Per soddisfare i requisiti di conformità aziendali e del settore (SEC Rule 17a-4(f)) per lo storage dei dati, ECS ha implementato quanto segue:

- **Rafforzamento della piattaforma:** il rafforzamento risolve le vulnerabilità di sicurezza in ECS, prevedendo misure quali il blocco della piattaforma per disabilitare l'accesso ai nodi o al cluster, la chiusura di tutte le porte non essenziali (ad esempio *ftpd, sshd*), la registrazione di audit completa dei comandi sudo e il supporto dei SRS (Dell EMC Secure Remote Services) per impedire l'accesso remoto ai nodi.
- **Report di conformità:** un agent di sistema segnala lo stato di conformità del sistema, ad esempio *Good* indica la conformità o *Bad* indica la non conformità.
- **Regole e retention dei record basati su policy:** possibilità di limitare le modifiche ai record o ai dati sottoposti a retention utilizzando policy, periodo di tempo e regole.
- **ARM (Advanced Retention Management):** per soddisfare i requisiti di conformità di Centera è stato definito un set di regole di retention solo per CAS.
  - **Retention basata su eventi:** abilita i periodi di retention che iniziano quando si verifica un evento specificato.
  - **Controversia:** abilita la prevenzione temporanea dell'eliminazione dei dati soggetti ad azioni legali.
  - **Min/max governor:** impostazione per singolo bucket del periodo di retention predefinito minimo e massimo.

La conformità viene abilitata a livello di namespace. I periodi di retention vengono configurati a livello di bucket. I requisiti di conformità certificano la piattaforma, ed è per questo motivo che la funzione di conformità è disponibile solo per ECS in esecuzione sull'hardware dell'appliance. Per informazioni sull'abilitazione e la configurazione della conformità in ECS, vedere la *Guida all'accesso ai dati di ECS* e la *Guida per l'amministratore di ECS* più recenti.

8 Implementazione

ECS può essere implementato come istanza di uno o più siti. Gli elementi di base di un deployment di ECS includono:

- **VDC (Virtual Data Center):** un cluster, noto anche come sito o area geograficamente distinta, costituito da un set di infrastruttura ECS gestito da una singola istanza di fabric.
- **Pool di storage (SP):** i pool di storage possono essere considerati come un sottoinsieme di nodi e lo storage associato appartenenti a un VDC. Un nodo può appartenere a un solo SP. EC è impostato a livello di SP con uno schema 12+4 o 10+2. Un SP può essere utilizzato come strumento per la separazione fisica dei dati tra client o gruppi di client che accedono allo storage su ECS.
- **Gruppo di replica (RG):** i gruppi di replica definiscono la posizione in cui è protetto il contenuto dei pool di storage e le posizioni da cui è possibile accedere ai dati. Un gruppo di replica con un unico sito di membri viene talvolta definito gruppo di replica locale. I dati sono sempre protetti in locale nella posizione in cui vengono scritti in caso di guasti del disco, del nodo e del rack. I gruppi di replica con due o più siti sono spesso definiti gruppi di replica globali. I gruppi di replica globali si estendono fino a 8 VDC e proteggono da guasti di dischi, nodi, rack e siti. Un VDC può appartenere a più gruppi di replica.
- **Namespace:** un namespace è concettualmente uguale a un tenant in ECS. Una caratteristica principale di un namespace consiste nel fatto che gli utenti di un namespace non possono accedere agli object in un altro namespace.
- **Bucket:** i bucket sono container per gli object creati in un namespace e talvolta considerati un container logico per i subtenant. In S3, i container sono denominati bucket e questo termine è stato adottato da ECS. In Atmos, l'equivalente di un bucket è un subtenant; in Swift l'equivalente di un bucket è un container, mentre per CAS un bucket è un pool di CAS. I bucket sono risorse globali in ECS. Ogni bucket viene creato in un namespace e ogni namespace viene creato in un gruppo di replica.

ECS sfrutta i seguenti sistemi di infrastruttura:

- **DNS:** (richiesto) ricerche in avanti e all'indietro necessarie per ciascun nodo di ECS.
- **NTP:** (richiesto) server Network Time Protocol.
- **SMTP:** (opzionale) server Simple Mail Transfer Protocol per l'invio di avvisi e report.
- **DHCP:** (opzionale) richiesto se vengono assegnati indirizzi IP tramite DHCP.
- **Fornitori di autenticazione:** (opzionale) gli amministratori di ECS possono essere autenticati utilizzando Active Directory e i gruppi LDAP. Gli utenti di object possono essere autenticati utilizzando Keystone. I provider di autenticazione non sono richiesti per ECS. ECS dispone di funzionalità di gestione degli utenti locali integrate, tuttavia gli utenti creati in locale non vengono replicati tra VDC.
- **Bilanciamento del carico:** (richiesto se lo impone il flusso di lavoro, altrimenti opzionale) il carico dei client deve essere distribuito tra i nodi per utilizzare in modo efficiente tutte le risorse disponibili nel sistema. Se è necessario un appliance o un servizio di bilanciamento del carico dedicato per gestire il carico tra i nodi di ECS, deve essere considerato un requisito. Gli sviluppatori che compilano applicazioni utilizzando l'SDK di ECS S3 possono sfruttare la funzionalità di bilanciamento del carico integrata. I servizi di bilanciamento del carico sofisticati possono tenere conto di ulteriori fattori, ad esempio il carico segnalato di un server, i tempi di risposta, lo stato up/down, il numero di connessioni attive e la posizione geografica. Il cliente è responsabile della gestione del traffico dei client e della determinazione dei requisiti di accesso. Indipendentemente dal metodo, sono generalmente prese in considerazione alcune opzioni di base, tra cui l'allocazione manuale dell'IP, il Round Robin DNS, il bilanciamento del carico sul lato client, gli appliance di bilanciamento del carico e i servizi di bilanciamento del carico geografici. Di seguito sono riportate brevi descrizioni di ciascuno di questi metodi:
  - **Allocazione manuale degli IP:** gli indirizzi IP vengono assegnati manualmente alle applicazioni. Questo non è generalmente raccomandato in quanto potrebbe non distribuire il carico o fornire tolleranza di errore.
  - **Round Robin DNS:** viene creata una voce del DNS che include tutti gli indirizzi IP del nodo. I client eseguono la query del DNS per risolvere gli FQDN per i servizi ECS e ricevono risposta con gli indirizzi IP di un nodo casuale. Ciò potrebbe fornire uno pseudo bilanciamento del carico. Questo metodo potrebbe non fornire la tolleranza di errore perché spesso viene utilizzato un intervento manuale per rimuovere gli indirizzi IP dei nodi non riusciti dal DNS. Con questo metodo possono

verificarsi problemi di Time To Live (TTL). Alcune implementazioni del server DNS possono memorizzare nella cache le ricerche DNS per un certo periodo in modo che i client che si connettono in un intervallo di tempo di chiusura possano essere associati allo stesso indirizzo IP, riducendo la quantità di distribuzione del carico ai nodi di dati. Non è consigliabile utilizzare DNS per distribuire il traffico in modalità Round Robin.

- **Bilanciamento del carico:** i servizi di bilanciamento del carico sono l'approccio più comune alla distribuzione del carico dei client. I client possono inviare il traffico a un servizio di bilanciamento del carico che lo riceve e lo inoltra a un nodo ECS integro. I controlli integrità proattivi o lo stato della connessione vengono usati per verificare la disponibilità di ogni nodo alle Service Request. I nodi non disponibili vengono rimossi dall'uso fino a quando non superano un controllo integrità. L'offload dell'elaborazione SSL con utilizzo intensivo della CPU può essere utilizzato per liberare tali risorse in ECS.
- **Bilanciamento del carico geografico:** il bilanciamento del carico geografico sfrutta DNS per instradare le ricerche a un appliance, come Riverbed SteelApp, che utilizza un Geo-IP o un altro meccanismo per determinare il sito migliore a cui instradare il client.

8.1 Deployment a un singolo sito

Durante un deployment iniziale a un singolo sito o un singolo cluster, i nodi vengono prima aggiunti a un SP. Gli SP sono container logici di nodi fisici. La configurazione di SP prevede la selezione del numero minimo richiesto di nodi disponibili e la scelta dello schema EC 12+4 predefinito oppure 10+2 con archiviazione a freddo. I livelli di avviso critici possono essere impostati inizialmente durante la configurazione dell'SP; in futuro, tuttavia, lo schema EC non può essere modificato dopo l'inizializzazione dell'SP. Il primo SP creato viene designato come SP di sistema e viene utilizzato per archiviare i metadati di sistema. L'SP di sistema non può essere eliminato.

I cluster contengono in genere uno o due SP, come illustrato nella Figura 21, uno per ogni schema EC, tuttavia se un'organizzazione richiede la separazione fisica dei dati, vengono utilizzati ulteriori SP per implementare i limiti.

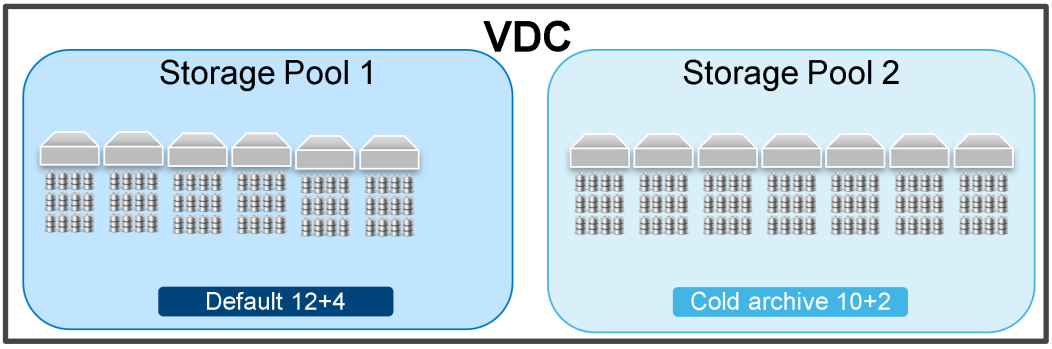

Figura 21 VDC con due pool di storage, ciascuno configurato con uno schema EC diverso

Dopo l'inizializzazione del primo SP, è possibile creare un VDC. La configurazione del VDC prevede la designazione degli endpoint di replica e gestione. Si noti che sebbene l'inizializzazione dell'SP di sistema sia richiesta prima della creazione del VDC, la configurazione del VDC non assegna gli SP ma gli indirizzi IP dei nodi.

Dopo aver creato un VDC, vengono configurati i gruppi di replica. I gruppi di replica sono risorse globali con configurazione che prevede la designazione di almeno un VDC nella configurazione del sito singolo o iniziale, insieme a uno degli SP del VDC. Un gruppo di replica con un singolo membro del VDC protegge i dati in locale a livello di disco, nodo e rack. La sezione successiva si espande nei gruppi di replica per includere i deployment di più siti.

I namespace sono risorse globali create e assegnate a un gruppo di replica. A livello di namespace vengono definite le policy di retention, le quote, la conformità e gli amministratori del namespace. ADO (Access During Outage) può essere configurato a livello di namespace, operazione descritta nella sezione successiva. In genere, è a livello di namespace che si organizzano i tenant. I tenant possono essere un'istanza dell'applicazione o un team, un utente, un gruppo di business o qualsiasi altro gruppo appropriato per l'organizzazione.

I bucket sono risorse globali e possono estendersi su più siti. La creazione di bucket comporta l'assegnazione a un namespace e a un gruppo di replica. Il livello bucket è dove viene abilitata la proprietà e l'accesso a file o CAS. La Figura 22 qui di seguito mostra un SP in un VDC con un namespace contenente due bucket.

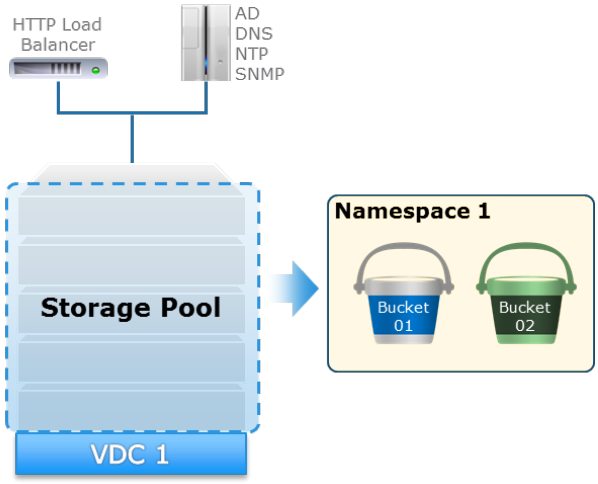

Figura 22 Esempio di deployment in un singolo sito

8.2 Deployment in più siti

Un deployment in più siti, definito anche ambiente federato o ECS federato, può estendersi su un massimo di otto VDC. I dati vengono replicati in ECS a livello di blocco. I nodi che partecipano a un gruppo di replica inviano i dati locali in modo asincrono a uno o a tutti gli altri siti. I dati vengono crittografati utilizzando AES256 prima di essere inviati attraverso la rete WAN tramite HTTP. I vantaggi principali riconosciuti durante la federazione di più VDC sono i seguenti:

- Consolidamento degli sforzi di gestione di più VDC in un'unica risorsa logica
- Protezione a livello di sito oltre a quella locale a livello di nodo, disco e rack
- Accesso geograficamente distribuito allo storage in modo ovunque attivo estremamente coerente

In questa sezione sul deployment in più siti vengono descritte le funzioni specifiche di ECS federato, ad esempio:

- **Coerenza dei dati:** per impostazione predefinita, ECS offre un servizio di storage estremamente coerente.
- **Gruppi di replica:** container globali utilizzati per designare i limiti di protezione e accesso.
- **Geocaching:** ottimizzazione dei flussi di lavoro per l'accesso al sito remoto nei deployment in più siti.
- **ADO:** comportamento dell'accesso ai client durante una situazione di interruzione temporanea dell'alimentazione del sito (TSO).

8.2.1 Coerenza dei dati

ECS è un sistema estremamente coerente che utilizza la proprietà per mantenere una versione autorevole di ciascun namespace, bucket e object. La proprietà viene assegnata al VDC in cui viene creato il namespace, il bucket o l'object. Ad esempio, se un namespace, NS1, viene creato in VDC1, VDC1 è proprietario di NS1 ed è responsabile della gestione della versione autorevole dei bucket all'interno di NS1. Se un bucket, B1 viene creato in VDC2 all'interno di NS1, VDC2 possiede B1 ed è responsabile della manutenzione della versione autorevole del contenuto del bucket e del VDC del proprietario di ogni object. Analogamente, se un object, O1 viene creato all'interno di B1 in VDC3, VDC3 possiede O1 ed è responsabile della gestione della versione autorevole di O1 e dei metadati associati.

La resilienza della protezione dei dati di più siti va a scapito dell'aumento dell'overhead di protezione dello storage e del consumo di larghezza di banda WAN. Le query di indice sono richieste quando si accede o si esegue l'aggiornamento di un object da un sito che non è proprietario dell'object. Analogamente, le ricerche di indici nella rete WAN sono richieste anche per recuperare informazioni, ad esempio un elenco autorevole di bucket in un namespace oppure object in un bucket, di proprietà di un sito remoto.

Comprendere in che modo ECS utilizza la proprietà per tenere traccia in modo autorevole dei dati a livello di namespace, bucket e object consente agli amministratori e ai proprietari delle applicazioni di prendere decisioni nella configurazione dell'ambiente per l'accesso.

8.2.2 Gruppo di replica attivo

Durante la creazione del gruppo di replica, è disponibile l'impostazione *Replicate to All Sites* che è disattivata per impostazione predefinita oppure può essere attivata per abilitare la relativa funzione. La replica dei dati in tutti i siti significa che i dati scritti singolarmente in ogni VDC vengono replicati in tutti gli altri VDC membri del gruppo di replica. Ad esempio, un'istanza ECS federata con un numero X di siti e un gruppo di replica attivo configurato per replicare i dati in tutti i siti comporterà un overhead di protezione X volte oppure un overhead di protezione dei dati X * 1,33 (o 1,2 nell'EC dell'archiviazione a freddo). La replica in tutti i siti può avere senso soprattutto per i data set più piccoli in cui l'accesso locale è importante. Se si disattiva questa impostazione, tutti i dati scritti in ogni VDC verranno replicati in un altro VDC. Il sito primario, in cui viene creato l'object, e il sito che archivia la copia di replica proteggono i dati in locale utilizzando lo schema EC assegnato all'SP locale. In altre parole, solo i dati originali vengono replicati attraverso la WAN e non tutti i segmenti di codifica EC associati.

I dati archiviati in un gruppo di replica attivo sono accessibili ai client tramite qualsiasi VDC membro del gruppo di replica disponibile. La Figura 23 qui di seguito illustra un esempio di un sistema ECS federato compilato con VDC1, VDC2 e VDC3. Vengono visualizzati due gruppi di replica: RG1 presenta un singolo membro, mentre VDC1 e RG2 hanno tutti e tre i VDC come membri. Vengono visualizzati tre bucket: B1, B2 e B3.

In questo esempio i client che accedono ai seguenti elementi:

- VDC1 ha accesso a tutti i bucket
- VDC2 e VDC3 hanno accesso solo ai bucket B2 e B3.

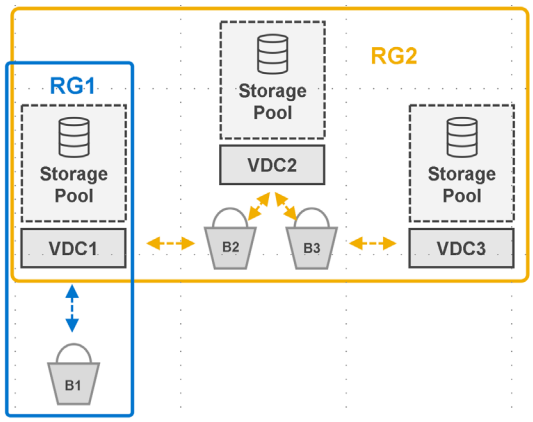

Figura 23 Accesso a livello di bucket per sito con gruppi di replica per uno o più siti

8.2.3 Gruppo di replica passivo

Un gruppo di replica passivo presenta tre VDC come membri. Due dei VDC sono designati come attivi e sono accessibili ai client. Il terzo VDC viene designato passivo e utilizzato solo come destinazione di replica. Il sito passivo viene utilizzato solo a scopo di ripristino e non consente l'accesso diretto ai client. I vantaggi della replica geo-passiva sono i seguenti:

- Riduzione dell'overhead di protezione dello storage aumentando il potenziale per le operazioni XOR
- Controllo a livello di amministratore della posizione utilizzata per lo storage di sola replica

Figura 24 La Figura 24 illustra un esempio di configurazione geo-passiva in cui VDC 1 e VDC 2 sono siti primari (source) che replicano i propri dati (blocchi) nella destinazione di replica, VDC 3.

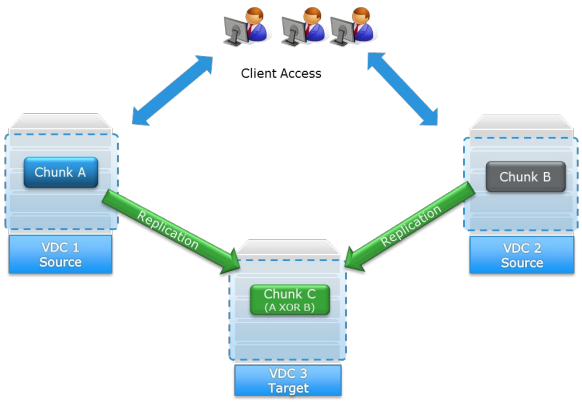

Figura 24 Percorsi di accesso ai client e di replica per il gruppo di replica geo-passivo

L'accesso di più siti a dati estremamente coerenti viene eseguito utilizzando la proprietà del namespace, dei bucket e degli object tra i siti membri del gruppo di replica. Le query di indice tra siti all'interno della WAN sono richieste quando l'accesso alle API deriva da un VDC che non è proprietario dei costrutti logici necessari. Le ricerche WAN vengono utilizzate per determinare la versione autorevole dei dati. Pertanto, se un object creato nel sito 1 viene letto dal sito 2, è necessaria una ricerca WAN per eseguire una query sul VDC proprietario dell'object, sito 1, per verificare se i dati dell'object replicati nel sito 2 sono la versione più recente dei dati. Se il sito 2 non dispone della versione più recente, recupera i dati necessari dal sito 1; in caso contrario, utilizza i dati replicati in precedenza. Questo processo è illustrato nella Figura 25 qui di seguito.

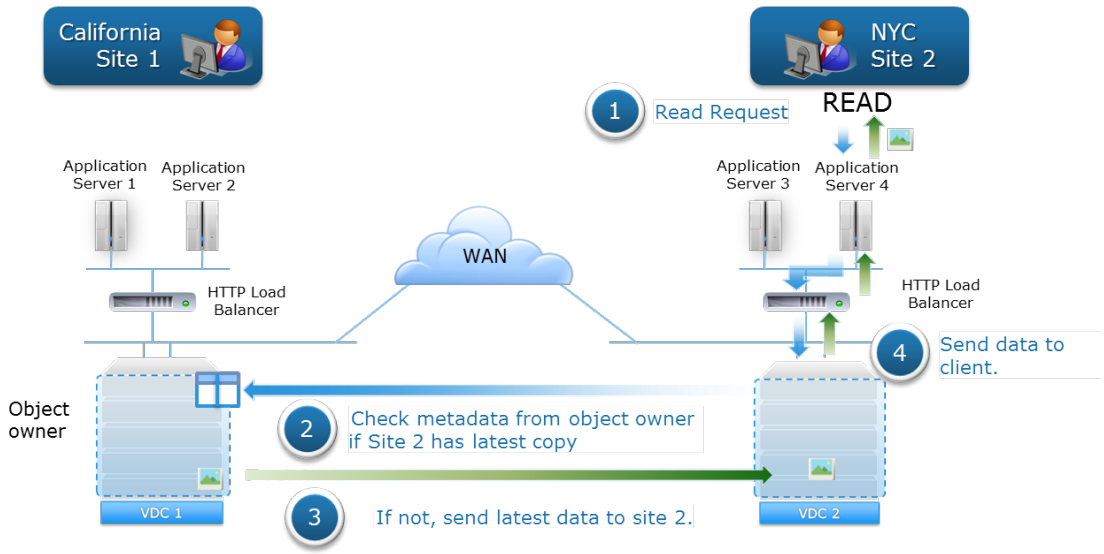

Figura 25 La richiesta di lettura a un VDC non proprietario attiva la ricerca WAN nel VDC proprietario dell'object

Il flusso di dati delle scritture in un ambiente con replica geografica in cui due siti aggiornano lo stesso object è illustrato nella Figura 26. In questo esempio, viene mostrato il sito 1 inizialmente creato che possiede l'object. L'object è stato codificato con erasure e le transazioni di journal correlate sono state scritte su disco nel sito 1. Il flusso di dati di un aggiornamento dell'object ricevuto nel sito 2 è il seguente:

1. Il sito 2 scrive innanzitutto i dati in locale.
2. Il sito 2 aggiorna in modo sincrono i metadati (scrittura nel journal) con il proprietario dell'object, sito 1, e attende l'aggiornamento dei metadati dal sito 1.
3. Il sito 1 riconosce la scrittura dei metadati nel sito 2.
4. Il sito 2 riconosce la scrittura nel client.

Nota: il sito 2 replica in modo asincrono i dati nel sito 1, il sito proprietario degli object, come di consueto. Se i dati devono essere serviti dal sito 1 prima di essere replicati dal sito 2, il sito 1 recupererà i dati direttamente dal sito 2.

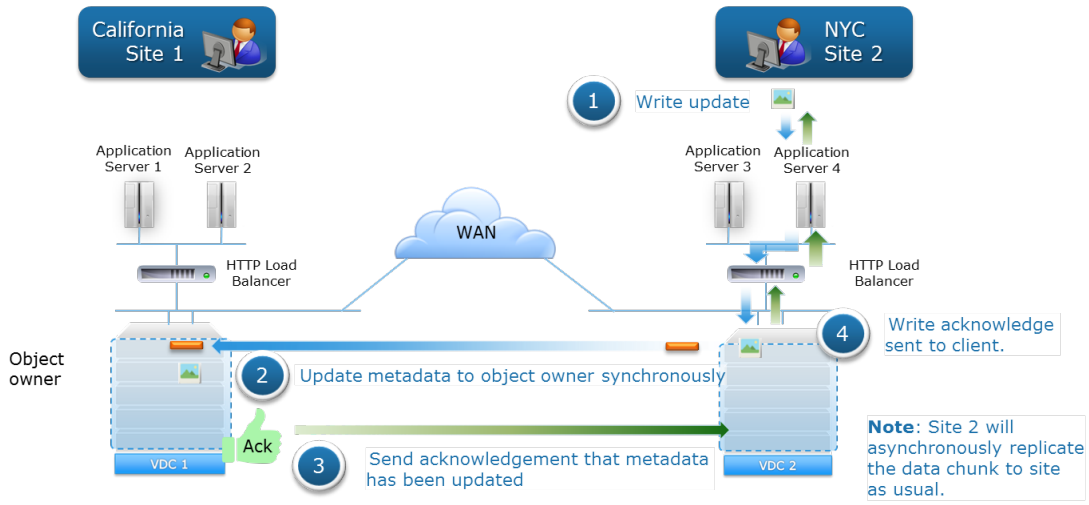

Figura 26 Aggiornamento dello stesso flusso di dati di object in un ambiente con replica geografica

Negli scenari di lettura e scrittura in un ambiente con replica geografica, è prevista una latenza per la lettura e l'aggiornamento dei metadati e il recupero dei dati dal sito proprietario degli object.

Nota: a partire da ECS 3.4, è possibile rimuovere un VDC da un gruppo di replica (RG) in una federazione multi-VDC senza influire sul VDC o su altri gruppi di replica associati al VDC. La rimozione di un VDC dal gruppo di replica non determina più un'interruzione permanente dell'alimentazione del sito (PSO). La rimozione di un VDC dal gruppo di replica avvia il ripristino.

Per ulteriori informazioni sul gruppo di replica, vedere la *Guida dell'amministratore di ECS* più recente

8.2.4 Memorizzazione di dati remoti nella cache geografica

ECS ottimizza i tempi di risposta per l'accesso ai dati archiviati nei siti remoti archiviando nella cache locale gli object letti sulla rete WAN. Ciò può essere utile per i modelli di accesso a più siti in cui i dati vengono spesso recuperati da un sito remoto o non proprietario. Si consideri un ambiente con replica geografica con tre siti, VDC1, VDC2 e VDC3, in cui un object viene scritto in VDC1 e la copia di replica dell'object viene archiviata in VDC2. In questo scenario, per soddisfare una richiesta di lettura ricevuta in VDC3 per l'object creato in VDC1 e replicato in VDC2, i dati dell'object devono essere inviati a VDC3 da VDC1 o VDC2. La memorizzazione nella cache geografica dei dati remoti a cui si accede di frequente consente di ridurre i tempi di risposta. Per la memorizzazione nella cache si fa ricorso a un algoritmo utilizzato meno di recente. La dimensione della cache geografica viene regolata quando l'infrastruttura hardware, ad esempio dischi, nodi e rack, viene aggiunta a un SP con replica geografica.

8.2.5 Comportamento durante l'interruzione dell'alimentazione del sito

L'interruzione temporanea dell'alimentazione del sito si riferisce in genere a un errore di connettività WAN o a un intero sito, ad esempio in caso di calamità naturali. TECS utilizza meccanismi heartbeat per rilevare e gestire gli guasti temporanei del sito. L'accesso client e la disponibilità dell'operazione API a livello di namespace, bucket e object durante un'interruzione temporanea dell'alimentazione del sito sono gestiti dalle seguenti opzioni ADO impostate a livello di namespace e bucket:

- **Disattivato (impostazione predefinita):** la coerenza elevata viene mantenuta durante un'interruzione temporanea dell'alimentazione.
- **Attivato:** alla fine è consentito un accesso coerente durante un'interruzione temporanea dell'alimentazione del sito.

La coerenza dei dati durante un'interruzione temporanea dell'alimentazione del sito viene implementata a livello di bucket. La configurazione avviene a livello di namespace e viene utilizzata l'impostazione ADO predefinita per ADO durante la creazione del nuovo bucket. Questa impostazione può esser cambiata al momento della creazione di un nuovo bucket; ciò che significa che il TSO può essere configurato per alcuni bucket e non per altri.

8.2.5.1 Accesso durante l'interruzione dell'alimentazione (ADO) non abilitata

Per impostazione predefinita, ADO non è abilitato e viene mantenuta una coerenza elevata. Tutte le richieste di API client in cui sono richiesti dati autorevoli relativi a namespace, bucket oppure object ma temporaneamente non disponibili avranno esito negativo. Le operazioni di lettura, creazione, aggiornamento ed eliminazione degli object, nonché i bucket di elenco non di proprietà di un sito online, avranno esito negativo. Anche le operazioni di creazione e modifica di bucket, utenti e namespace avranno esito negativo.

Come accennato in precedenza, il proprietario del sito iniziale di bucket, namespace e object è il sito in cui la risorsa è stata creata per la prima volta. Durante un'interruzione temporanea dell'alimentazione del sito, alcune operazioni potrebbero non riuscire se il proprietario del sito della risorsa non è accessibile. I punti salienti delle operazioni consentite o non consentite durante un'interruzione temporanea dell'alimentazione del sito includono:

- La creazione, l'eliminazione e l'aggiornamento di bucket, namespace, utenti di object, provider di autenticazione, mapping di utenti e gruppi NFS non sono consentiti da alcun sito.
- Elencare i bucket all'interno di un namespace è consentito se il sito del proprietario del namespace è disponibile.

HDFS/NFS consente ai bucket di proprietà del sito inaccessibile di essere read-only.

8.2.5.2 ADO abilitato

In un bucket abilitato per ADO, durante un'interruzione temporanea dell'alimentazione del sito, il servizio di storage fornisce risposte coerenti. In questo scenario le letture e, facoltativamente, le scritture da un sito secondario (non proprietario) vengono accettate e rispettate. Inoltre, una scrittura in un sito secondario durante un'interruzione temporanea dell'alimentazione del sito fa sì che il sito secondario assuma la proprietà dell'object. Ciò consente a ogni VDC di continuare a leggere e scrivere object dai bucket in un namespace condiviso. Infine, la nuova versione dell'object diventa la versione autorevole dell'object durante la riconciliazione successiva all'interruzione temporanea dell'alimentazione del sito anche se un'altra applicazione aggiorna l'object nel VDC proprietario.

Sebbene molte operazioni di object continuino durante un'interruzione dell'alimentazione della rete, alcune operazioni non sono consentite, ad esempio la creazione di nuovi bucket, namespace o utenti. Quando viene ripristinata la connettività di rete tra due VDC, il meccanismo heartbeat rileva automaticamente la connettività, ripristina il servizio e riconcilia gli object provenienti dai due VDC. Se lo stesso object viene aggiornato in VDC A e VDC B, la copia nel VDC non proprietario è la copia autorevole. Pertanto, se un object di proprietà di VDC B viene aggiornato sia su VDC A che VDC B durante la sincronizzazione, la copia nel VDC A sarà la copia autorevole che viene mantenuta e verrà annullato il riferimento all'altra copia, che diventerà disponibile per il recupero dello spazio.

Quando più di due VDC fanno parte di un gruppo di replica e se la connettività di rete viene interrotta tra un VDC e gli altri due, le operazioni di scrittura/aggiornamento/proprietà continuano come nel caso di due VDC. Tuttavia, il processo di risposta alle richieste di lettura è più complesso, come descritto di seguito.

Se un'applicazione richiede un object di proprietà di un VDC non raggiungibile, ECS invia la richiesta al VDC con la copia secondaria dell'object. Tuttavia, la copia del sito secondario potrebbe essere stata soggetta a un'operazione di contrazione dei dati, ovvero uno XOR tra due data set diversi che produce un nuovo data set. Pertanto, il VDC del sito secondario deve prima recuperare i blocchi dell'object inclusi nell'operazione XOR originale e deve eseguire l'operazione XOR di tali blocchi con la copia di ripristino. Questa operazione restituirà il contenuto del blocco originariamente archiviato nel VDC non riuscito. I blocchi dell'object ripristinato possono quindi essere riassemblati e restituiti. Quando i blocchi vengono ricostruiti, vengono anche memorizzati nella cache in modo che il VDC possa rispondere più rapidamente alle richieste successive. Nota: la ricostruzione richiede molto tempo. Più VDC in un gruppo di replica implicano più blocchi che devono essere recuperati da altri VDC e quindi la ricostruzione dell'object richiede più tempo.

Se si verifica un guasto irreparabile, un intero VDC potrebbe non essere più recuperabile. ECS tratta il VDC irrecuperabile come guasto del sito temporaneo. Se il guasto è permanente, il System Administrator deve eseguire il failover permanente del VDC dalla federazione per avviare l'elaborazione del failover, che avvia la risincronizzazione e la riprotezione degli object archiviati nel VDC non riuscito. Le attività di ripristino vengono eseguite come processo in background. È possibile rivedere lo stato di avanzamento del ripristino nel portale ECS.

Un'opzione bucket aggiuntiva è disponibile per ADO *read-only (RO)* che garantisce che la proprietà dell'object non venga mai modificata e rimuove la possibilità di conflitti altrimenti causati dagli aggiornamenti degli object nei siti non riusciti e online durante un'interruzione temporanea dell'alimentazione del sito. Lo svantaggio di ADO RO è che durante un'interruzione temporanea dell'alimentazione del sito non è possibile creare nuovi object e nessun object esistente nel bucket può essere aggiornato fino a quando tutti i siti non sono tornati online. L'opzione ADO RO è disponibile solo durante la creazione del bucket, non può essere modificata in seguito. Per impostazione predefinita, l'opzione è disabilitata.

Tabella 8 Tolleranza ai guasti di più siti

Modello di guasto	Tolleranza
Ambiente con replica geografica	Fino a un guasto del sito

8.3 Tolleranza ai guasti

ECS è progettato per tollerare una gamma di situazioni di guasto delle apparecchiature utilizzando una serie di domini di guasti. La gamma di condizioni di guasto si estende su un ambito variabile, tra cui:

- Guasto di singolo disco rigido in un singolo nodo
- Guasto di più dischi rigidi in un singolo nodo
- Più nodi con guasto di singolo disco rigido
- Più nodi con guasto di più dischi rigidi
- Guasto di un singolo nodo
- Gusto di più nodi
- Perdita di comunicazione con un VDC replicato
- Perdita di un intero VDC replicato

In una configurazione con un sito, due siti o con replica geografica, l'impatto del guasto dipende dalla quantità e dal tipo di componenti interessati. Tuttavia, a ogni livello, ECS fornisce meccanismi per difendersi dall'impatto dei guasti dei componenti. Molti di questi meccanismi sono già stati trattati in questo documento, ma qui e nella Figura 27 vengono ripresi per mostrare come vengono applicati alla soluzione. tra cui:

- Guasto del disco
  - Segmenti EC o copie di replica dallo stesso blocco non vengono archiviati nello stesso disco
  - Calcolo del checksum sulle operazioni di scrittura e lettura
  - Verifica della coerenza in background che verifica nuovamente i checksum

- Guasto del nodo
  - Distribuire segmenti o copie di replica di un blocco equamente tra i nodi in un VDC
  - ECS Fabric mantiene i servizi in esecuzione e gestisce risorse quali dischi e rete.
  - Record e tabelle di partizione protetti dal failover della proprietà della partizione da nodo a nodo.
- Guasto del rack nel VDC
  - Distribuire segmenti o copie di replica di un blocco equamente tra i rack in un VDC
  - Un'istanza del registro del fabric viene eseguita in ogni rack e può essere riavviata in qualsiasi altro nodo nello stesso rack in caso di errore del nodo.

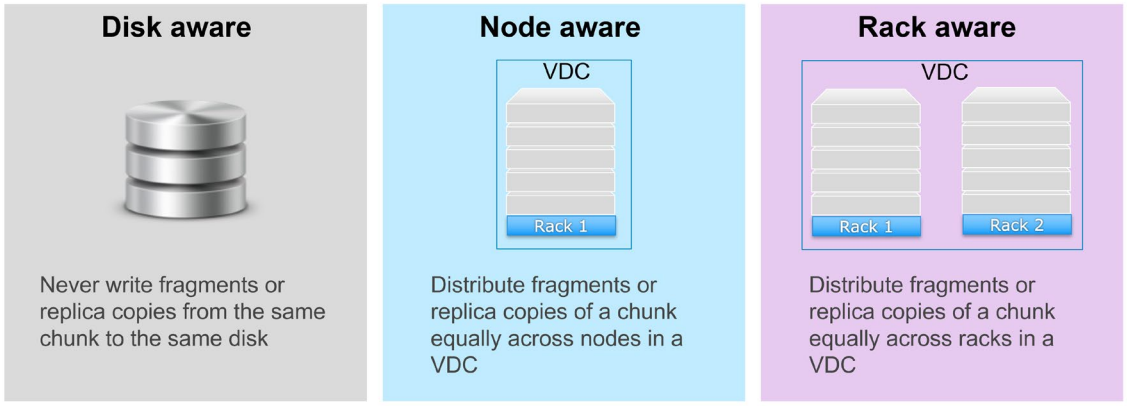

Figura 27 Meccanismi di protezione a livello di disco, nodo e rack

Il grafico seguente definisce il tipo e il numero di guasti dei componenti da cui ciascuno schema EC fornisce protezione per la configurazione di base dei rack. La Tabella 9 Protezione con codifica di erasure codice tra domini di guasti evidenzia l'importanza di considerare l'impatto dei domini di guasti protettivi sulla disponibilità complessiva dei dati e dei servizi in termini di numero di nodi richiesti in ciascuno schema EC.

Tabella 9 Protezione con codifica di erasure codice tra domini di guasti

Schema EC	N. di nodi in VDC	N. di frammenti di blocchi per nodo	Dati EC protetti da...
12 + 4 Predefinito	5 o meno	4	• Perdita fino a quattro dischi o • Perdita di un nodo
	6 o 7	3	• Perdita fino a quattro dischi o • Perdita di un nodo e di un disco da un secondo nodo
	8 o più	2	• Perdita fino a quattro dischi o • Perdita di due nodi o • Perdita di un nodo e due dischi
	16 o più	1	• Perdita di quattro nodi o • Perdita di tre nodi e dischi da un nodo aggiuntivo o • Perdita di due nodi e dischi da un massimo di due nodi diversi o • Perdita di un nodo e dischi da un massimo di tre nodi diversi o • Perdita di quattro dischi da quattro nodi diversi
10+2 Storage non attivo	11 o meno	2	• Perdita fino a due dischi o • Perdita di un nodo
	12 o più	1	• Perdita di un numero qualsiasi di dischi da due nodi diversi o • Perdita di due nodi

8.4 Automazione della sostituzione dei dischi

A partire da ECS 3.5, i clienti possono sostituire i dischi guasti con Dell EMC Services utilizzando un intuitivo flusso di lavoro del portale ECS (UI web). Questa funzione offre:

- Risoluzione fai da te dei guasti delle unità
- Riduzione del tempo di riparazione dei guasti
- Flessibilità operativa e risparmi sul TCO

La pagina di manutenzione nel portale ECS fornisce agli amministratori visibilità su tutti i dischi di ciascun nodo. In caso di guasto di un'unità, il sistema avvia automaticamente il ripristino. Tutti i tipi di risorse sull'unità vengono ripristinati e quando l'unità è pronta per essere rimossa dal nodo, il portale ECS visualizza il pulsante di sostituzione, come illustrato nella Figura 28.

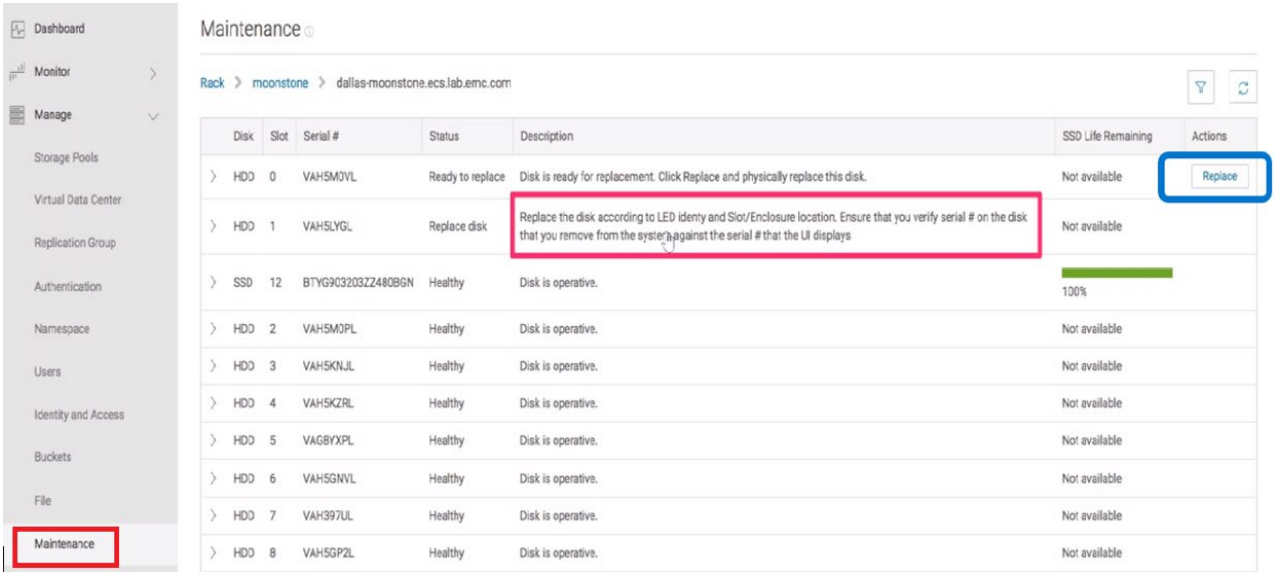

Figura 28 Automazione della sostituzione dei dischi

Nota: è necessario rimuovere una sola unità alla volta. In questo modo si evita di sostituire l'unità sbagliata.

8.5 Tech Refresh

Il Tech Refresh è una funzione gestita direttamente da Dell EMC Professional Services, disponibile a partire da ECS 3.5, che consente di rimuovere senza interruzioni i nodi hardware meno recenti dai cluster ECS utilizzando la funzionalità software integrata. Si tratta di un'operazione efficiente e a basso utilizzo di risorse che può essere scaglionata con precisione. Questa funzione riduce l'overhead prima associato alla disattivazione di hardware ECS.

Il Tech Refresh include tre parti:

- **Estensione dei nodi:** aggiunta di nodi Gen3 al cluster esistente
- **Migrazione delle risorse:** trasferimento di tutte le risorse dai nodi esistenti ai nodi Gen 3
- **Evacuazione dei nodi:** pulizia dei nodi meno recenti e loro rimozione dal cluster

Professional Services deve essere coinvolto durante la manutenzione Tech Refresh. Per ulteriori informazioni su Tech Refresh, vedere la Guida al Tech Refresh di *ECS.*

9 Overhead della protezione dello storage

Ogni membro del VDC in un gruppo di replica è responsabile della propria protezione EC dei dati a livello locale. In altre parole, i dati vengono replicati, ma non i segmenti di codifica correlati. Sebbene EC sia più efficiente rispetto ad altre forme di protezione, ad esempio il mirroring completo delle unità di copia, comporta un overhead intrinseco dei costi di storage a livello locale. Tuttavia, quando è necessario disporre di copie secondarie replicate off-site e tutti i siti che hanno accesso ai dati quando un singolo sito diventa non disponibile, i costi di storage diventano più alti rispetto a quando si utilizzano i tradizionali metodi di protezione della copia dei dati da sito a sito. Ciò è particolarmente vero quando i dati univoci vengono distribuiti fra tre o più siti.

ECS fornisce un meccanismo in cui l'efficienza dell'overhead di protezione dello storage può aumentare con la federazione di tre o più siti. In un ambiente replicato con due VDC, ECS replica i blocchi dal VDC primario o proprietario in un sito remoto per garantire high availability e resilienza. Non è possibile eludere il costo del 100% dell'overhead di protezione di una copia completa dei dati in un deployment ECS federato con due siti.

Si considerino ora tre VDC in un ambiente con più siti, VDC1, VDC2 e VDC3, in cui ogni VDC dispone di dati univoci replicati da ognuno degli altri VDC. VDC2 e VDC3 possono inviare una copia dei dati a VDC1 per la protezione. VDC1 avrebbe quindi i propri dati originali, oltre a replicare i dati da VDC2 e VDC3. Ciò significa che VDC1 archivierebbe 3 volte la quantità di dati scritti nel proprio sito.

In questo caso, ECS può eseguire un'operazione XOR dei dati VDC2 e VDC3 archiviati in locale in VDC1. Questa operazione matematica confronta quantità uguali di dati univoci, blocchi ed esegue il rendering di un nuovo blocco che contiene un numero sufficiente di caratteristiche dei due blocchi di dati originali per rendere possibile il ripristino di uno dei due set originali. Quindi, dove in precedenza c'erano tre set univoci di blocchi di dati archiviati in VDC1, consumando 3 volte la capacità disponibile, ora ne sono presenti solo due: il data set locale originale e le copie di protezione ridotte XOR.

In questo stesso scenario, se VDC3 non è più disponibile, ECS può ricostruire i blocchi di dati di VDC3 utilizzando le copie dei blocchi richiamate da VDC2 e i dati (C1 ⊕C2) di VDC3 archiviati in locale in VDC1. Questo principio si applica a tutti e tre i siti che partecipano al gruppo di replica e dipende da ciascuno dei tre VDC che presentano data set univoci. La Figura 29 illustra un calcolo XOR con due siti che eseguono la replica in un terzo sito.

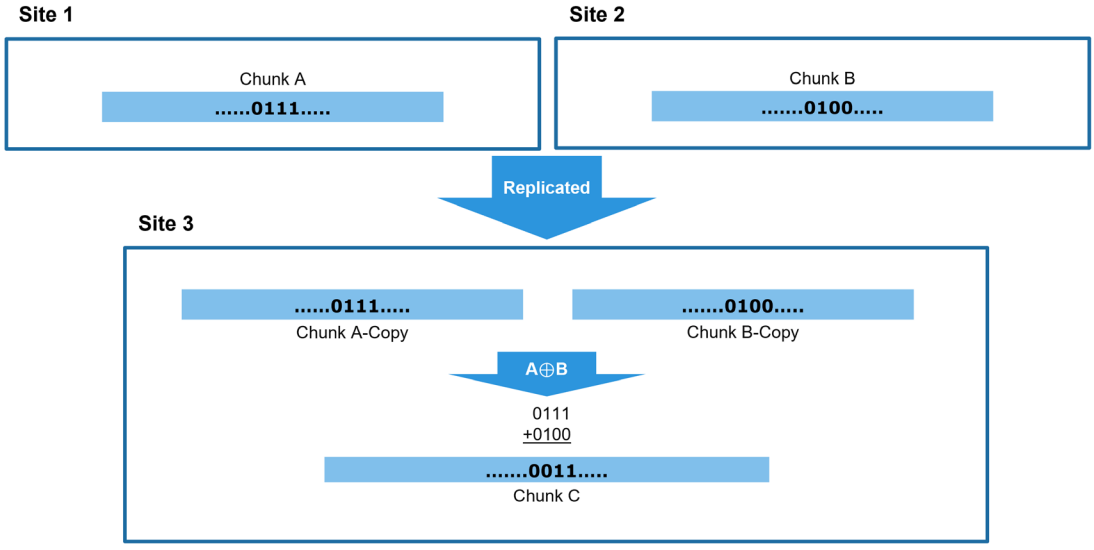

Figura 29 Efficienza nella protezione dei dati XOR

Se accordi sui livelli di servizio aziendali richiedono velocità di accesso in lettura ottimali anche in caso di guasto del sito completo, l'impostazione di replica in tutti i siti impone a ECS di ripristinare le copie complete dei dati replicati da archiviare in tutti i siti. Ciò fa aumentare i costi di storage in proporzione al numero di VDC che partecipano al gruppo di replica. Pertanto una configurazione di 3 siti ripristinerebbe l'overhead di protezione dello storage 3X. L'impostazione Replicate to All Sites è disponibile durante la creazione del gruppo di replica e non può essere attivata e disattivata a piacere.

Con l'aumentare del numero di siti federati, l'ottimizzazione XOR è più efficiente nel ridurre l'overhead di protezione dello storage dovuto alla replica. La Tabella 10 fornisce informazioni sull'overhead della protezione dello storage in base al numero di siti per un normale EC di 12+4 e un EC di archiviazione a freddo di 10+2, illustrando come ECS può diventare più efficiente in termini di storage man mano che vengono collegati più siti.

Nota: per ridurre l'overhead dei dati replicati in tre e fino a otto siti, i dati univoci devono essere scritti in modo relativamente uguale in ciascun sito. Scrivendo i dati in quantità uguali tra i siti, ogni sito avrà un numero simile di blocchi di replica. Numeri simili di blocchi di replica in ogni sito portano a un numero simile di operazioni XOR che possono verificarsi in ogni sito. L'efficienza massima dello storage di più siti si ottiene riducendo il numero massimo di blocchi di replica archiviati tramite XOR.

Tabella 10 Overhead della protezione dello storage

N. di siti nel gruppo di replica	EC 12+4	EC 10+2
1	1,33	1.2
2	2,67	2.4
3	2.00	1,8
4	1,77	1,6
5	1,67	1,5
6	1.60	1.44
7	1,55	1.40
8 (n. di siti max nel gruppo di replica)	1.52	1.37

10 Conclusioni

Le organizzazioni si trovano ad affrontare costi sempre crescenti per dati e storage, in particolare nello spazio del cloud pubblico. L'architettura con scalabilità orizzontale e distribuzione geografica di ECS offre una piattaforma cloud on-premise che scala nell'ordine di exabyte di dati con un *costo totale di proprietà* significativamente inferiore rispetto al cloud storage pubblico. ECS è un'ottima soluzione per la sua versatilità, iperscalabilità, le sue funzioni avanzate e l'uso di hardware di uso comune.

A Supporto tecnico e risorse

http://www.dell.com/support Dell.com/support è concentrata sul soddisfare le esigenze dei clienti con servizi e supporto comprovati.

I documenti e i video tecnici di storage forniscono competenze che contribuiscono a garantire il successo dei clienti sulle piattaforme di storage Dell EMC.