

## WHITE PAPER

---

# Backup e ripristino: maggiore efficienza e riduzione dei costi IT con la deduplicazione dei dati

Sponsorizzato da: EMC Corporation

---

Laura DuBois

Robert Amatruda

Febbraio 2010

## INTRODUZIONE

La deduplicazione dei dati sta migliorando in modo significativo lo scenario economico dell'IT grazie a requisiti meno complessi associati allo spazio richiesto dallo storage, a finestre di backup ridotte e a un minor consumo della larghezza di banda di rete sia negli ambienti distribuiti che nei data center. All'interno di ambienti reali, la deduplicazione consente di aumentare l'efficienza dei processi di backup e ripristino, riducendo al contempo i costi IT. Questo white paper esamina i vari approcci alla deduplicazione per il backup dei dati e indica quali sono gli elementi da considerare nella scelta di una soluzione. Illustra inoltre le soluzioni di backup e ripristino incluse nel portafoglio EMC di prodotti per la deduplicazione e i relativi casi di utilizzo per garantire backup più efficienti e costi ridotti.

---

## Adozione della tecnologia di deduplicazione

Attualmente la domanda per la tecnologia di deduplicazione dei dati in ambienti di medie e grandi dimensioni registra un costante aumento: le aziende, infatti, sono alla continua ricerca di metodi per stare al passo con la crescita dello storage, che raddoppia ogni anno. Tale crescita è dovuta principalmente a fattori quali l'implementazione di nuove applicazioni, la proliferazione di sistemi virtualizzati, la creazione di archivi di documenti elettronici, la condivisione dei contenuti, l'utilizzo delle tecnologie Web 2.0 e la conservazione dei record digitali. In considerazione dei budget IT limitati, la necessità di contenere la crescita dei dati è ancora più evidente dal momento che le aziende cercano di ridurre i costi operativi e di capitale. Dal punto di vista fisico, molti responsabili di data center si trovano a dover gestire infrastrutture con restrizioni in termini di consumo, raffreddamento e ingombro. La deduplicazione è una tecnologia che consente non solo di ottimizzare l'efficienza dello storage riducendo i costi, ma di risolvere anche le problematiche associate ai data center con restrizioni fisiche.

Con la deduplicazione è anche possibile superare i problemi dovuti alle inefficienze a livello di gestione, backup e rete. Poiché i dati sono in continuo aumento, sempre più sproporzionato appare il rapporto tra il numero di tecnici IT e i volumi di storage da gestire. La deduplicazione riduce lo spazio occupato dai dati, riequilibrando questo rapporto. Analogamente, con l'aumento del divario tra la potenza di elaborazione del server e il disco, le aziende devono cercare di migliorare le prestazioni degli ambienti su rete WAN, all'interno dei sottosistemi di storage su disco e nelle finestre di backup limitate. La tecnologia di deduplicazione dei dati può anche ottimizzare l'infrastruttura fisica e virtuale esistente trasferendo una quantità di dati inferiore attraverso collegamenti di rete locali o remoti, migliorare i tempi di risposta ai livelli di servizio e ridurre le finestre di backup. Infine, grazie all'utilizzo di supporti ad accesso casuale (disco), la

deduplicazione migliora i tempi di ripristino, la sicurezza dei dati e il livello di affidabilità.

Recentemente la virtualizzazione ha posto qualche problema. Infatti, dal momento che le aziende continuano a implementare tecnologie basate su VM a supporto del consolidamento dei server e del disaster recovery, i sistemi virtuali elaborano dati che possono essere altamente ridondanti e che richiedono comunque protezione. Per gestire i vari guasti o ripristinare un'immagine, è in genere necessario integrare un server fisico e file discreti in una singola soluzione o in un processo di backup. La deduplicazione offre un significativo risparmio della capacità di storage in quanto consente di eliminare la ridondanza tipica dei file VMDK. Tuttavia, approcci standard quali l'implementazione di un agente di backup tradizionale in un sistema guest virtuale e l'utilizzo di un proxy VCB per creare un backup a livello di immagine non contribuiscono a ridurre il volume dei dati delle macchine virtuali da sottoporre a backup oppure i requisiti di larghezza di banda della rete per lo spostamento dei dati. In combinazione con un software di backup, questa tecnologia assicura inoltre una protezione completa, conveniente ed efficiente degli ambienti basati su sistemi virtuali.

---

## I vantaggi della deduplicazione

Le aziende implementano la deduplicazione dei dati in alcuni punti dello stack dell'infrastruttura per risolvere una serie di problematiche. I vantaggi offerti dalla deduplicazione sono:

- ☒ **Riduzione dei costi:** la deduplicazione ottimizza l'efficienza delle risorse, offrendo al contempo un risparmio sui costi associato alla riduzione dei requisiti di consumo, raffreddamento e ingombro del data center e alla razionalizzazione della capacità di storage, della larghezza di banda della rete e dello staff IT.
- ☒ **Miglioramento dei livelli di servizio per il backup e il ripristino:** la deduplicazione può migliorare significativamente le prestazioni di backup, in modo da ridurre i tempi di esecuzione. Sfruttando lo storage su disco ad accesso casuale, assicura inoltre prestazioni di ripristino superiori rispetto a quelle dei metodi ad accesso sequenziale (nastro).
- ☒ **Maggiore convenienza del disco rispetto al nastro:** la deduplicazione consente l'esecuzione del backup su disco per più applicazioni. Il nastro viene ancora utilizzato nei data center aziendali per la sua convenienza e le funzionalità di archiviazione offerte. Tuttavia, la riduzione del costo per GB in caso di utilizzo del disco insieme alla deduplicazione rende i relativi costi uguali o inferiori rispetto al nastro.
- ☒ **Riduzione delle emissioni di anidride carbonica:** questa tecnologia riduce non solo i requisiti di storage in termini di consumo, raffreddamento e ingombro, ma anche le emissioni di anidride carbonica, per una maggiore responsabilità ambientale.

La tecnologia di deduplicazione consente di risolvere molti dei problemi che affliggono, da oltre un decennio, le aziende di piccole e grandi dimensioni, tra cui la crescita dei dati (che raddoppiano ogni anno), il rispetto di finestre di backup più brevi, l'esecuzione di ripristini rapidi in caso di guasti operativi e così via.

Nella Tabella 1 sono riportati alcuni dei principali problemi di backup oltre a una breve descrizione di come è possibile risolverli implementando la deduplicazione.

**TABELLA 1**

Problemi di backup e impatto della deduplicazione

<b>Problematiche relative al backup</b>	<b>Impatto della deduplicazione</b>
Con l'esecuzione di attività 24x7, le <b>finestre di backup</b> sono troppo brevi per soddisfare i requisiti generali dei clienti.	I backup tradizionali comportano il trasferimento di grandi quantità di dati ridondanti, che possono superare finestre di backup ridotte o inesistenti. Grazie alla deduplicazione integrata nel software di backup è possibile ridurre il volume di dati da sottoporre a backup; inoltre, i sistemi di storage con deduplicazione rapida in linea possono migliorare le performance della destinazione di backup, consentendo così di eseguire il backup di una maggiore quantità di dati entro la finestra disponibile.
I <b>tempi di ripristino</b> stanno diventando troppo brevi per consentire una riduzione dei costi associati all'inattività.	La deduplicazione riduce i costi relativi allo storage su disco di un maggior numero di dati di backup. Con il backup su disco anziché su nastro è possibile migliorare significativamente i tempi di ripristino di numerose applicazioni.
La scarsa <b>affidabilità dei backup</b> mette a rischio il ripristino dei dati.	L'esecuzione di backup su nastro implica una serie di rischi, tra cui guasti ai supporti (ad es. supporti e testine danneggiati), esaurimento dei supporti disponibili o guasti hardware. Dal momento che la deduplicazione utilizza il disco nel processo di protezione dei dati, le possibilità di guasto sono inesistenti o ridotte al minimo. L'utilizzo del disco semplifica inoltre l'esecuzione di verifiche dello stato e di altre funzionalità di risoluzione automatica degli errori o di prevenzione dai guasti.
La <b>maggiore virtualizzazione dei server</b> riduce le risorse disponibili per il backup, con un conseguente aumento dei tempi per la relativa esecuzione e il superamento delle finestre.	La deduplicazione può essere utilizzata per eliminare l'elaborazione delle risorse condivise associate ai dati ridondanti, ridurre i conflitti per le risorse fisiche e velocizzare i backup delle macchine virtuali. Consente inoltre di conservare più a lungo i dati di backup delle macchine virtuali occupando molto meno spazio, con la garanzia che il ripristino operativo avvenga rapidamente dal disco e mai dal nastro.
A causa della <b>crescita dei dati</b> non è possibile eseguire il backup di tutti i dati nelle finestre disponibili.	La quantità di dati aziendali da proteggere aumenta del 50% ogni anno. Questa crescita è in contrasto con le finestre di backup notturne limitate e con i metodi tradizionali. La deduplicazione è in grado di supportare l'aumento dei dati, garantendo al contempo backup efficienti di dataset estesi.
La <b>copia fuori sede</b> con metodi a nastro tradizionali espone i dati al rischio di perdita o furto.	In caso di guasto, l'utilizzo di nastri rimovibili per lo storage fuori sede rischia di compromettere il supporto fisico. In combinazione con processi di replica sicuri, la deduplicazione consente di conservare una copia elettronica fuori sede, eliminando la necessità di gestire manualmente i supporti a nastro e migliorando quindi la sicurezza.
I <b>dati distribuiti in filiali remote</b> devono essere protetti e ripristinati a livello centralizzato.	Le filiali remote stanno sostituendo i processi standalone di backup su nastro con un approccio di backup centralizzato di tipo <i>edge-to-core</i> in grado di migliorare il backup, il ripristino e la gestione. La deduplicazione rende possibile l'invio di grandi volumi di dati di backup su collegamenti WAN congestionati a un data center.
I <b>costi per l'infrastruttura di backup</b> stanno aumentando per tenere il passo con la crescita della capacità e le finestre di backup.	Per le aziende, l'implementazione di più infrastrutture a nastro si traduce in una serie di problematiche a livello di crescita dei dati e finestre di backup. Se da un lato l'aggiunta di unità a nastro e l'automazione eliminano i colli di bottiglia prestazionali consentendo di eseguire più rapidamente i backup, dall'altro aumentano i costi e il sovraccarico di gestione. La deduplicazione risolve il problema alla radice in modo che sia possibile ridurre i costi dell'infrastruttura a nastro supportando al contempo sia la crescita della capacità che le finestre di backup.

Fonte: IDC, 2010

## **DEDUPLICAZIONE: COSA, DOVE, COME E QUANDO**

---

### ***Cos'è la deduplicazione***

Secondo IDC la deduplicazione dei dati è una tecnologia che standardizza i dati duplicati su un singolo oggetto dati condiviso per garantire efficienza in termini di capacità di storage. Nello specifico, fa riferimento a un qualsiasi algoritmo che ricerca dati duplicati (ad es. blocchi, gruppi e segmenti) ed elimina quelli doppi dopo averli individuati. I dati duplicati rilevati non vengono infatti conservati. Affinché il sistema di storage possa fare riferimento a una copia esatta dell'oggetto dati già memorizzato su disco, il "puntatore dei dati" viene modificato di conseguenza. Inoltre, la deduplicazione dei dati riduce i costi relativi alla conservazione di più copie dello stesso oggetto dati.

La deduplicazione dei dati viene spesso associata ai processi di confronto dei sottofile e si distingue dallo storage a singola istanza (SIS), che invece confronta i dati a livello di file o di oggetto. La deduplicazione dei sottofile analizza un file e lo suddivide in "segmenti" più piccoli, che vengono valutati in presenza di contenuti di dati ridondanti su più sistemi e postazioni. La deduplicazione si distingue anche dalla compressione, in grado di ridurre lo spazio occupato da un singolo oggetto anziché quello occupato da più file o porzioni di file. I dati deduplicati, inoltre, possono anche essere compressi in modo da ridurne ulteriormente l'ingombro.

### ***Dove avviene la deduplicazione***

La deduplicazione dei dati per il backup può avvenire all'origine o sulla destinazione. Un esempio di deduplicazione all'origine consiste nel ridurre le dimensioni dei dati di backup sul client (ad es. Exchange o un file server) in modo che solo i dati di sottofile univoci vengano inviati sulla rete LAN o WAN durante il processo di backup. Nel caso della deduplicazione all'origine, l'applicazione di backup integra la tecnologia di deduplicazione nella sua architettura. Con la deduplicazione sulla destinazione, ad esempio, vengono invece ridotte le dimensioni dei dati di backup sulla rete locale che hanno raggiunto un sistema di storage di deduplicazione. La deduplicazione all'origine fornisce vantaggi in termini di larghezza di banda della rete LAN e WAN, finestre di backup e storage. Nel caso della deduplicazione sulla destinazione, il sistema di storage integra la tecnologia di deduplicazione nel relativo controller. Oltre ad essere compatibile con il software di backup esistente, la deduplicazione sulla destinazione consente di ridurre i costi di storage e l'impatto della replica sulla rete WAN. Negli ambienti in cui è installata la tecnologia di deduplicazione non solo si ottengono diversi vantaggi, ma si riducono anche i tempi e i costi di implementazione. Le aziende devono valutare i propri problemi di backup e associarli ai vari approcci di deduplicazione.

#### ***Deduplicazione all'origine***

La deduplicazione all'origine (o sul software di backup del client) offre una serie di vantaggi che vanno oltre l'ottimizzazione della capacità di storage. Consente infatti di trasferire una minore quantità di dati dal dispositivo di origine al repository di storage, con una conseguente decongestione dell'infrastruttura virtuale o fisica e dei collegamenti LAN/WAN. Poiché vengono inviati solo

i segmenti di dati di sottofile nuovi o modificati, il volume di dati spostati si riduce notevolmente, consentendo così l'esecuzione di backup completi giornalieri in tempi rapidi. Sebbene il sovraccarico incrementale sulla CPU del client possa raggiungere livelli del 15% in caso di deduplicazione all'origine, il backup viene completato in meno tempo rispetto ai metodi convenzionali; alcune architetture, inoltre, offrono funzionalità di regolazione per la gestione degli aumenti temporanei del sovraccarico. In un periodo di sette giorni l'impatto della deduplicazione all'origine è notevolmente inferiore rispetto a quello generato dagli agenti tradizionali. Inoltre, la deduplicazione all'origine è flessibile da implementare: gli uffici remoti più piccoli, infatti, possono implementare semplicemente l'agente di backup del software. Per gli ambienti con database di dimensioni molto grandi o database con elevate percentuali di modifica giornaliera potrebbe rivelarsi più adatta una soluzione di deduplicazione sulla destinazione. Fortunatamente i fornitori ricorrono a strumenti di valutazione specifici per aiutare le aziende a scegliere la soluzione più adatta.

### ***Deduplicazione sulla destinazione***

La deduplicazione sulla destinazione ottimizza la capacità di storage su disco per il backup in quanto vengono archiviati su disco solo i dati di sottofile nuovi e univoci. Tutti i dati di backup vengono tuttavia inviati ancora alla destinazione di deduplicazione utilizzando un tradizionale software di backup per garantire integrazione con l'infrastruttura IT esistente. Questo consente di rispettare le finestre di backup disponibili solo se era la destinazione di backup precedente, in genere il nastro, a rappresentare il collo di bottiglia per le prestazioni della soluzione di backup. Con la deduplicazione sulla destinazione, il sistema di storage stesso (chiamato anche sistema di storage per la deduplicazione) esegue la deduplicazione al fine di ottimizzare la protezione dei dati e le prestazioni di disaster recovery, evitando così che il processo venga eseguito dagli application server. La deduplicazione sulla destinazione si implementa facilmente creando un sistema di storage veloce e indipendente dall'applicazione (collegabile come lo storage NAS su Ethernet o una libreria VTL su Fibre Channel) che non richiede l'installazione di software per client o altre configurazioni. I sistemi di storage per la deduplicazione vengono spesso utilizzati con data set e database di grandi dimensioni. La deduplicazione sulla destinazione può anche essere utilizzata nei data center centrali per grandi volumi di dati e in postazioni remote per il backup locale seguito dalla replica su un data center centrale.

---

### ***Quando eseguire la deduplicazione***

Attualmente esistono due diversi approcci relativi al momento di esecuzione della deduplicazione: la modalità in linea o quella post-processo. La deduplicazione in linea elimina i dati ridondanti prima della loro scrittura su disco e non richiede quindi un'area temporanea, mentre quella post-processo analizza e riduce i dati dopo averli archiviati su disco richiedendo, quindi, un'area temporanea a piena capacità sulla quale avviare un processo di deduplicazione. Nella scelta dell'approccio più adatto le aziende devono valutare aspetti quali la velocità di backup e la capacità su disco.

La deduplicazione in linea è un metodo di deduplicazione più efficiente ed economico. Consente di ridurre notevolmente la capacità raw su disco richiesta dal sistema dal momento che non viene mai scritto su disco il dataset completo non ancora deduplicato. Se nell'ambito del processo di deduplicazione in linea

è supportata la replica, questo approccio riduce anche i tempi di disaster recovery poiché il sistema non deve attendere per assorbire l'intero dataset e deduplicarlo prima che abbia inizio la replica sul sito remoto.

Con la deduplicazione post-processo è invece necessario aspettare che i dati vengano trasferiti su disco prima di eseguire la deduplicazione. Questo approccio, che richiede una capacità iniziale superiore rispetto al metodo in linea, genera sfasamenti prima del completamento della deduplicazione e della replica. Vi è inoltre il rischio di incoerenza tra un sistema locale e un sistema remoto dal momento che esistono due zone di storage, ciascuna con regole e comportamenti diversi da gestire.

---

### ***Come viene eseguita la deduplicazione***

Le modalità di deduplicazione variano in base al tipo di implementazione. La deduplicazione basata su hash suddivide un flusso di file o di backup in più blocchi di dati di sottofile a lunghezza fissa o variabile. Per ogni segmento viene calcolato un valore hash. Questo processo calcola un valore univoco per ciascun segmento, che viene registrato in un indice. Se un file è aggiornato vengono salvati solo i dati di sottofile modificati; la modifica non richiede il salvataggio di un nuovo file. È importante fare una distinzione tra le implementazioni basate su hash con segmenti a lunghezza fissa e quelle con segmenti a lunghezza variabile. Adottando un approccio a lunghezza variabile è possibile regolare – in modo dinamico e in base al contenuto – le dimensioni di un segmento per “ospitare” segmenti di dati ridondanti la cui posizione è stata modificata o compensata in un flusso di byte mentre si modifica un file. Un approccio a lunghezza fissa non riconosce i dati ridondanti riposizionati o compensati; di conseguenza esegue nuovamente e in modo inefficiente il backup dei segmenti in quanto questi sembrano univoci, anche se in realtà sono già presenti nel repository di backup. L'indice hash viene tenuto in memoria, ma quando aumentano gli elementi in esso contenuti può espandersi dalla memoria al disco, richiedendo agli I/O del disco di ricercare i valori hash. I fornitori offrono diverse soluzioni per consentire alle aziende di risolvere questi problemi pratici associati alla tecnologia: il risultato può quindi andare dall'eliminazione del problema a un peggioramento significativo delle prestazioni.

In alternativa, è possibile ricorrere alla deduplicazione dei dati basata su delta (nota anche come differenziazione delta o crittografia delta), che si occupa di archiviare o trasmettere i dati sotto forma di differenze rispetto a una copia di riferimento. La copia di riferimento è una copia completa di dati utilizzata per ricreare altre versioni dei dati stessi. La deduplicazione dei dati basata su delta può essere eseguita a livello di blocco o di byte. Anziché utilizzare un hash per determinare i nuovi dati, il metodo basato sulla differenziazione delta scansiona e indicizza il flusso dei dati in entrata in modo da individuare i dati simili a quelli già archiviati. Poiché non è necessario elaborare un hash “forte”, tra i vantaggi offerti da un approccio basato su delta vi è anche il minore utilizzo della CPU. Tuttavia, un processo di differenziazione delta richiede numerosi I/O del disco per confrontare i vecchi dati con i nuovi dati in entrata. Il vantaggio di ogni singolo approccio, nel lungo periodo, può quindi dipendere dai miglioramenti nelle prestazioni della CPU in funzione della tecnologia a disco implementata.

Un altro fattore che può incidere sul rapporto di deduplicazione è rappresentato dalla capacità del motore di deduplicazione di riconoscere i marcatori inseriti nel

flusso di dati dall'applicazione di backup o un particolare formato dei dati (ad esempio un'applicazione di backup, dati di Microsoft Exchange, ecc.). Affinché si possano riconoscere i marcatori e il formato dei dati è necessario sapere il punto in cui i metadati specifici dell'applicazione vengono integrati in un flusso. Se il motore di deduplicazione è in grado di riconoscere i marcatori o il formato dei dati, le dimensioni del segmento possono essere adattate in base al formato dei dati dell'applicazione originaria, per risultati di deduplicazione potenzialmente migliori. Per poter utilizzare questo approccio è tuttavia necessario comprendere i formati variabili di ciascuna applicazione di backup (NetWorker, NBU, TSM, ecc.) o applicazione utente (Oracle, Exchange, ecc.).

## CONSIDERAZIONI PER LA SCELTA DELLA TECNOLOGIA DI DEDUPLICAZIONE

Attualmente sul mercato sono disponibili diversi tipi di prodotti che offrono funzionalità di deduplicazione. Anche le applicazioni di backup, le appliance, le librerie VTL, le soluzioni di ottimizzazione della WAN e i sottosistemi primari di storage su disco possono includerne alcune. Per un'azienda è importante definire quali sono i problemi da risolvere sulla base dell'applicazione o del tipo di dati prima di scegliere l'opzione di deduplicazione più adatta. Ogni singolo approccio alla deduplicazione offre, infatti, vantaggi diversi in termini di capacità, prestazioni ed efficienza di rete.

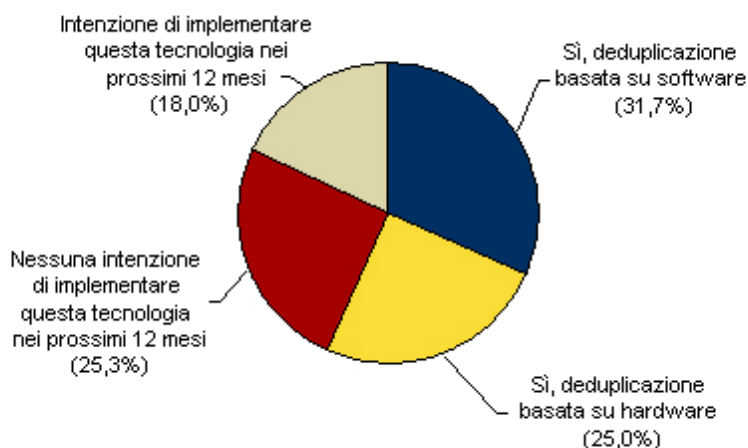
1. **Rapporti di deduplicazione:** il rapporto di deduplicazione ottenuto varia in base a una serie di fattori, inclusi il tipo e la percentuale di variazione dei dati, il periodo di conservazione, il tipo di segmenti (a lunghezza variabile o fissa), le regole di backup, la capacità di riconoscimento del formato dei file e così via. Dagli studi IDC emerge che, sulla base dei fattori sopracitati, i rapporti di deduplicazione associati allo storage su disco back-end in ambienti reali sono compresi tra 8:1 e 22:1. Le soluzioni di deduplicazione all'origine consentono di ridurre notevolmente la larghezza di banda della rete richiesta ogni giorno rispetto ai tradizionali metodi di backup completo. Tuttavia, come per tutti i parametri relativi alle prestazioni, i valori variano a seconda dell'ambiente. Ai fini della scelta, le aziende devono tenere in considerazione il throughput, la scalabilità e le prestazioni offerte, nonché testare la deduplicazione nei propri ambienti e con i dataset aziendali.
2. **Compressione, crittografia e multiplexing:** la compressione, ovvero la codifica dei dati per ridurre lo spazio che occupano, può essere implementata come tecnologia complementare alla deduplicazione. La compressione è ottimizzata per singoli oggetti e ne riduce l'ingombro, mentre la deduplicazione "lavora" sugli oggetti. Tuttavia, la compressione può essere eseguita sui dati che sono già stati duplicati, così da ridurre ulteriormente lo spazio occupato. Se la deduplicazione viene applicata a un file già compresso o crittografato, i vantaggi offerti da questa tecnologia sono minimi, se non inesistenti (a meno che il file compresso non venga nuovamente sottoposto a backup). Le aziende che utilizzano la deduplicazione insieme alla compressione possono ottenere ulteriori vantaggi, ma solo se la deduplicazione viene eseguita prima. È anche importante considerare l'utilizzo attuale del multiplexing per il backup, che interallaccia i dati da più client in un unico flusso inviato a un'unità a nastro. Questo processo rende tuttavia difficile il rilevamento dei segmenti di dati esistenti. Per sfruttare al massimo i vantaggi derivanti dalla deduplicazione

occorre disattivare il multiplexing, una funzionalità presente in numerose applicazioni di backup che viene utilizzata per evitare il *shoeshining* e migliorare le prestazioni di scrittura dei dati sul supporto a nastro. Tuttavia, in questo scenario, la disattivazione del multiplexing non annulla i miglioramenti delle prestazioni di backup ottenuti con la sua applicazione.

3. **Deduplicazione per macchine virtuali:** l'utilizzo di macchine virtuali nell'ambiente di produzione ha aumentato la necessità di proteggere e ripristinare le VM, l'host fisico e i file. Le opzioni per il backup di macchine virtuali includono il backup a livello di file e/o immagine attraverso l'utilizzo di un agente di backup in esecuzione su un sistema guest o di una console di servizio oppure di un'interfaccia API di backup del fornitore delle soluzioni di virtualizzazione. Le soluzioni di backup tradizionali non sono adatte per il backup di macchine virtuali in quanto trasferiscono grandi quantità di dati ridondanti e richiedono molti cicli della CPU per l'esecuzione di un backup, con conseguenti scarse prestazioni di backup e minore consolidamento del server. La deduplicazione consente di risolvere anche queste problematiche. Con un approccio di deduplicazione all'origine, i dati duplicati non vengono mai trasmessi sull'infrastruttura fisica sottostante condivisa, al fine di garantire backup completi rapidi ed efficienti. La deduplicazione può essere eseguita anche a livello globale, su file VMDK, per eliminare il backup dei dati ridondanti all'interno dei sistemi virtuali. Ciò migliora notevolmente il ripristino delle macchine virtuali senza nastro garantendo al contempo un disaster recovery efficace grazie alle funzionalità di replica dei sistemi di deduplicazione.
4. **Deduplicazione per uffici di filiali remote:** come per le operazioni di data center, anche per gli uffici remoti è necessario eseguire sia il ripristino locale che il disaster recovery (in remoto). Tuttavia, le caratteristiche degli uffici remoti comportano alcuni problemi come, ad esempio, la limitata larghezza di banda della WAN, l'assenza di personale IT dedicato e il numero sproporzionato di data center regionali o principali. La deduplicazione consente di ridurre al minimo lo spostamento dei dati sulla WAN e di eliminare i dati ridondanti a livello di filiale e data center. Disponendo di pochi tecnici IT presso gli uffici remoti, molte aziende sono alla ricerca di metodi in grado di ridurre l'ingombro dell'hardware di storage sulle postazioni distribuite. Per limitare questo problema, la deduplicazione all'origine può essere implementata tramite software. Nei casi in cui occorre eseguire rapidamente il ripristino in locale, è possibile inoltre implementare sistemi di deduplicazione più piccoli sui siti remoti; alcuni fornitori supportano la replica da questi sistemi a un data center centralizzato. In un recente studio, IDC ha analizzato l'utilizzo della deduplicazione nelle postazioni di uffici remoti e il tipo di tecnologia implementata (vedere Figura 1).

**FIGURA 1**

Utilizzo della tecnologia di deduplicazione per la protezione dei dati nelle filiali remote



n = 300

Fonte: IDC's Remote Branch Special Study, 2009

5. **Deduplicazione per data center di produzione/disaster recovery:** le aziende hanno ancora difficoltà a soddisfare le finestre di backup per alcune applicazioni all'interno dei data center di grandi dimensioni e parallelamente non possono permettersi un peggioramento delle prestazioni di backup. Questo è lo scenario ideale per implementare un approccio alla deduplicazione in grado di includere la deduplicazione all'origine e sulla destinazione a seconda del tipo di applicazione e dell'ambiente. Ottimizzare la larghezza di banda della rete all'interno di un data center può essere meno urgente che eseguire la replica remota su un sito di disaster recovery. Tuttavia, dal momento che le finestre di backup continuano a ridursi, col tempo la larghezza di banda della rete diventerà un problema.
6. **Deduplicazione e replica:** la replica è la prossima sfida che la tecnologia di deduplicazione dovrà affrontare. I fornitori presenti da tempo sul mercato non solo hanno dimostrato che funziona se viene progettata e implementata correttamente, ma confermano anche le valutazioni positive espresse dagli utenti e la crescente domanda. La deduplicazione viene implementata negli ambienti aziendali – periferici e centrali – aumentando l'efficienza e riducendo al contempo i costi dell'infrastruttura. La replica remota è ormai diventata di fondamentale importanza dal momento che sempre più aziende se da un lato riducono al minimo l'utilizzo del nastro nelle postazioni remote, dall'altro necessitano ancora del nastro in una postazione centralizzata a scopo di archiviazione e conformità. I requisiti degli utenti relativi alla replica sono sempre più complessi e includono:
  - ❑ **Replica con riconoscimento della deduplicazione per replicare dataset deduplicati e non volumi completi:** alcuni fornitori offrono servizi di replica integrati nelle soluzioni per la deduplicazione.

Le aziende devono assicurarsi che le funzionalità di replica riconoscano la deduplicazione.

- ❑ **Replica completa e a livello di directory/nastro:** alcune applicazioni consentono l'esecuzione di una replica completa del sistema, mentre altre richiedono la flessibilità necessaria a determinare le condivisioni o i nastri virtuali da replicare.
- ❑ **Monitoraggio delle repliche, tuning delle prestazioni e risoluzione dei problemi:** nonostante sia stata implementata la tecnologia di deduplicazione, la maggior parte delle aziende di grandi dimensioni necessita ancora di replicare grandi quantità di dati. Questi dati vengono gestiti sulla base di un processo di replica asincrona o pianificata, che controlla l'esecuzione delle repliche e la larghezza di banda utilizzata. Grazie a strumenti specifici per il tuning delle prestazioni e la risoluzione dei problemi è possibile garantire un processo di replica coerente nei tempi previsti.
- ❑ **Replica pianificata e in tempo reale per collegamenti a latenza più bassa o elevata:** alcuni collegamenti/siti consentono l'esecuzione della replica in tempo reale, mentre altri richiedono un processo di replica pianificato. Gli uffici delle filiali remote presentano caratteristiche molto diverse tra loro e possono disporre di collegamenti a latenza inferiore, problema che potrebbe non interessare i collegamenti tra due data center.

7. **Seeding e migrazione:** la deduplicazione, pur essendo ideale per ridurre lo storage e/o lo spostamento dei dati ridondanti, richiede la definizione di una baseline iniziale o l'esecuzione di un primo backup. Per la deduplicazione e la replica *edge-to-core* gli utenti devono valutare come definire questa base di riferimento su collegamenti con larghezza di banda limitata. Molti fornitori offrono servizi di seeding di diverso tipo per creare rapidamente questa baseline tramite processi di replica in massa con riconoscimento della deduplicazione sulla base di sistemi side-by-side oppure utilizzando una serie di nastri dell'ultimo backup completo e ripristinandoli in locale su un sistema di deduplicazione. Se si eseguono cicli di aggiornamento dello storage con rotazione di 3-5 anni, è anche importante considerare le modalità di esecuzione della migrazione e valutare al contempo gli eventuali tempi di interruzione per l'ambiente esistente.
8. **Scelta del fornitore:** ogni fornitore rilascia dichiarazioni sulla validità dell'approccio alla deduplicazione offerto. Da una ricerca IDC emerge tuttavia che non tutte le soluzioni di deduplicazione attualmente sul mercato funzionano come dichiarato dal produttore. Nella scelta della soluzione di deduplicazione più adatta le aziende devono tener conto di alcuni fattori come, ad esempio, da quanto tempo viene commercializzato il prodotto, quanti clienti lo utilizzano nell'ambiente di produzione e qual è la sua efficacia negli ambienti aziendali reali. È importante inoltre valutare il livello di scalabilità di un prodotto. A questo proposito occorre richiedere riferimenti per l'utente e una matrice di supporto del sistema e/o dell'applicazione. Le aziende che decidono di non eseguire una prova di fattibilità possono avere brutte sorprese associate al mancato soddisfacimento dei requisiti di prestazioni e affidabilità.
9. **Casi di utilizzo della deduplicazione:** la deduplicazione è una tecnologia che consentirà allo stack dell'infrastruttura di storage di evolversi

ulteriormente. Attualmente questa tecnologia è stata ampiamente implementata negli ambienti di backup con un grande volume di dati ridondanti. Questi dati vengono sottoposti a backup ogni settimana, con un conseguente utilizzo non necessario del server, della rete e delle risorse di storage. Alcune aziende stanno valutando o testando soluzioni di deduplicazione esistenti negli ambienti di storage principali sulla base di un approccio NAS. Questa implementazione richiede un miglioramento delle prestazioni in modo da evitare implicazioni legate alla latenza e ai tempi di risposta. Attualmente la tecnologia di deduplicazione è ideale per il backup di macchine virtuali, per gli uffici remoti e le filiali, oltre che per gli ambienti di data center.

## **PORTAFOGLIO EMC DI SOLUZIONI DI DEDUPLICAZIONE**

EMC offre un'ampia gamma di soluzioni e servizi per il backup e il ripristino che consentono alle aziende di ridurre i costi dell'IT ed eseguire backup più efficienti. Tra le soluzioni di backup con funzionalità di deduplicazione sono inclusi il software EMC Avamar, il sistema di storage EMC Data Domain ed EMC NetWorker, che può essere implementato con Avamar, Data Domain e altri sistemi di destinazione di terze parti. Sebbene non siano descritte in questo white paper, EMC offre una soluzione di deduplicazione per lo storage primario e i dati di backup con il sistema di storage NAS EMC Celerra e una soluzione di deduplicazione per l'archiviazione su disco nell'ambito della gamma di prodotti Centera.

---

### **EMC Avamar**

EMC Avamar è un software di backup in grado di integrare la tecnologia di deduplicazione per individuare i dati ridondanti all'origine e ridurre al contempo la quantità di informazioni da sottoporre al backup prima del loro trasferimento su reti LAN o WAN. Avamar consente alle aziende di ridurre il numero di dati ed eseguire backup completi giornalieri in ambienti VMware, uffici remoti, desktop/laptop e server LAN e NAS all'interno di data center. Offre, inoltre, funzionalità di deduplicazione dei dati di backup non solo su più siti e server, ma anche nel tempo. Diversamente dai prodotti basati su metodi di ripristino tradizionali, Avamar ripristina rapidamente i dati in un unico passaggio, eliminando la necessità di ripristinare l'ultimo backup completo valido e i relativi backup incrementali per raggiungere il punto di ripristino desiderato. Le funzionalità offerte da questa soluzione si distinguono in modo significativo da quelle delle tradizionali soluzioni di backup.

Oltre a tenere traccia dei file nuovi o modificati, l'agente Avamar identifica questi dati senza analizzare l'intera struttura del file system ricercandoli dapprima nella cache dei file in locale. Una volta identificati, l'agente suddivide i file nuovi o modificati in segmenti dati di sottofile a lunghezza variabile assegnando un valore hash (ID univoco) a ciascun segmento. Verifica quindi se l'hash è già presente nella cache hash locale e se è stato precedentemente sottoposto a backup. Nel qual caso il backup non viene rieseguito. Infine, l'agente comunica con il server Avamar per verificare se il valore hash è univoco oppure già presente. I segmenti di dati nuovi vengono trasferiti sulla rete LAN o WAN mentre è in corso il backup completo giornaliero.

Rispetto agli agenti di backup tradizionali, questi processi aumentano l'utilizzo della CPU sull'host. Inoltre, dal momento che il backup protegge efficacemente solo i segmenti di dati nuovi, i backup Avamar vengono eseguiti in minor tempo rispetto ai tradizionali backup completi e incrementali. Ad esempio, un backup incrementale generalmente eseguito in 10 ore può essere completato in circa un'ora utilizzando Avamar. In questo modo è possibile ridurre l'impatto settimanale dei backup da 50 a 5 ore per incrementi da lunedì a venerdì. Anche i backup completi giornalieri eseguiti con Avamar sono notevolmente più veloci rispetto a quelli tradizionali.

Avamar offre ora anche la protezione per desktop e laptop, che viene eseguita in background utilizzando i collegamenti di rete esistenti senza pregiudicare i cicli della CPU. Il backup dei dati utente viene eseguito automaticamente al momento dell'accesso entro le normali finestre di backup; in alternativa, l'utente può avviare il ripristino dei dati su richiesta.

Le soluzioni di backup e ripristino Avamar consentono la deduplicazione all'origine e globale, funzionalità che le rendono ideali per le aziende che intendono:

- Implementare macchine virtuali e valutare nuove strategie di protezione per il ripristino di server fisici, server virtuali e oggetti discreti
- Migliorare i backup di uffici in filiali remote per eseguire rapidamente backup giornalieri, centralizzare la gestione, ottimizzare l'affidabilità, garantire processi di replica sicuri e ridurre il traffico di backup sui collegamenti WAN congestionati
- Contenere la crescita dei dati, soddisfare le finestre di backup e gestire il traffico di rete per il backup di ambienti NAS e file server locali
- Proteggere i dati di desktop e laptop, inclusi quelli negli uffici e sui dispositivi mobili

Le implementazioni di EMC Avamar possono essere configurate come:

- Opzioni client
  - Gli agenti del software Avamar vengono implementati sul sistema da proteggere (client), senza che si renda necessario aggiungere ulteriori prodotti hardware in locale (ad es. media server). Gli agenti sono già presenti per la maggior parte dei sistemi operativi e per le applicazioni e i principali database. A differenza di molte applicazioni di backup, il software client di Avamar non è a pagamento in quanto sfrutta il modello di licensing basato sulla capacità deduplicata. Le implementazioni risultano quindi estremamente convenienti, con un conseguente ampliamento della base clienti.
- Opzioni server (repository di backup)
  - Server Avamar di terze parti: il software Avamar può essere acquistato e implementato su una serie di server certificati standard di settore con storage su disco interno.

- ❑ Avamar Data Store: questa soluzione *all-in-one* include il software Avamar preinstallato e preconfigurato sull'hardware EMC per semplificare gli ordini e le operazioni di distribuzione e assistenza.
- ❑ Avamar Virtual Edition per VMware: questa configurazione, la prima del settore nel suo genere, consente di implementare il server Avamar come dispositivo virtuale sul server ESX esistente, sfruttando le risorse e lo storage su disco disponibili.

Avamar è diverso dagli altri prodotti per la deduplicazione all'origine disponibili oggi sul mercato. Ad esempio, la deduplicazione Avamar utilizza segmenti dati di sottofile a lunghezza variabile in grado di assicurare una maggiore efficienza e prestazioni superiori. Avamar sfrutta anche un'architettura grid per garantire prestazioni e capacità scalabili, all'interno della quale ogni nodo incrementale ottimizza la CPU, la memoria, gli I/O e lo storage per l'intero sistema.

Il grid Avamar si basa su una configurazione RAIN (Redundant Array of Independent Nodes) in grado di garantire fault tolerance integrata e high availability sull'intero grid, eliminando al contempo ogni SPF. Avamar distribuisce l'indice generato internamente sui suoi nodi per assicurare affidabilità, bilanciamento del carico e scalabilità. Inoltre, verifica quotidianamente e in modo automatico che i dati di backup siano completamente ripristinabili; il server Avamar esegue invece due volte al giorno un controllo su se stesso per verificare la propria integrità. Infine, Avamar supporta numerose applicazioni e vari client, inclusi Exchange, SQL, Oracle, DB2, SharePoint, Lotus Notes e il protocollo NDMP.

Avamar offre diversi metodi di protezione dei sistemi virtuali e di quelli fisici. Le opzioni per il backup Avamar di ambienti di macchine virtuali VMware includono:

- ☒ **Agente Avamar nel sistema operativo guest:** l'integrazione di un agente Avamar all'interno di ciascun sistema operativo guest assicura una maggiore efficienza rispetto agli approcci basati sui tradizionali agenti di backup. Gli agenti Avamar in versione light consentono di ridurre non solo il numero dei dati sul sistema guest, ma anche i requisiti di rete e le contese di risorse condivise a livello di CPU, scheda NIC, disco e memoria. Dal momento che vengono sottoposti a backup solo i dati di sottofile nuovi o univoci, Avamar permette di eseguire rapidamente backup giornalieri completi.
- ☒ **Avamar per il backup VCB o per il backup di API vStorage:** un agente Avamar in esecuzione su un server proxy esegue solo il backup dei dati univoci e ripartisce l'elaborazione per i sistemi guest. La deduplicazione, che avviene all'interno e sui file VMDK, supporta il backup sia a livello di file che di immagine. Grazie alle efficienti funzionalità di replica offerte da Avamar è possibile trasferire rapidamente i file VMDK sulla rete WAN in modo da soddisfare gli obiettivi di disaster recovery.
- ☒ **Agente Avamar sulla console ESX:** con un agente Avamar sulla console ESX è possibile eseguire la deduplicazione all'interno e su file VMDK. Questo metodo offre funzionalità di backup e ripristino a livello di immagine indipendenti da VMware VCB o dallo storage condiviso. Non consente tuttavia il ripristino a livello di file.

## EMC Data Domain

I sistemi di storage con funzionalità di deduplicazione Data Domain consentono di ridurre la capacità di storage su disco necessaria per la conservazione e la protezione dei dati aziendali. Individuando i dati e i file ridondanti al momento della loro archiviazione, i sistemi Data Domain riducono mediamente da 10 a 30 volte lo spazio occupato dallo storage rispetto ai data set originari. I dati di backup possono quindi essere replicati e recuperati sulle reti esistenti per operazioni di disaster recovery e consolidamento su nastro più semplici.

I sistemi Data Domain sono disponibili in numerose configurazioni con prestazioni e capacità diverse. Molte opzioni software migliorano la funzionalità dei relativi sistemi, ottimizzandone il valore. Ad esempio:

- ☒ Data Domain Appliance Series offre una serie di sistemi scalabili con storage integrato e funzionalità di deduplicazione convenienti e ad elevato throughput.
- ☒ Data Domain DDX Array Series offre un sistema di storage scalabile e ad alte prestazioni dotato di array di dischi con capacità logica fino a 56,7 PB e un throughput fino a 86,4 TB/ora. Se si utilizzano i controller DD880, questo sistema è in grado di gestire e supportare un massimo di 2.880 postazioni remote.
- ☒ Data Domain Gateway Series offre gateway di classe enterprise con funzionalità di deduplicazione e compressione per i data center che utilizzano sistemi di storage esterni di terze parti.
- ☒ Data Domain Replicator Software è un'efficace soluzione software in rete con replica automatizzata per il disaster recovery, la protezione di dati in uffici remoti e il consolidamento del nastro su più siti.
- ☒ Data Domain Virtual Tape Library Software emula più librerie a nastro su un'interfaccia Fibre Channel per fornire storage con deduplicazione negli ambienti SAN.
- ☒ Data Domain OpenStorage Software consente la perfetta integrazione tra i sistemi di storage con funzionalità di deduplicazione Data Domain e Symantec Veritas NetBackup.
- ☒ Data Domain Retention Lock Software consente agli utenti di implementare facilmente la deduplicazione con blocco dei file in modo da soddisfare i requisiti di conformità e governance IT. Gli amministratori IT possono quindi disporre della flessibilità operativa necessaria per eseguire le attività aziendali quotidiane al costo giusto. Il software include inoltre una serie di funzionalità per la distruzione sicura dei dati.

Le soluzioni di deduplicazione Data Domain forniscono un livello di storage consolidato per dati di backup, nearline e archivio. Inoltre, si integrano con l'infrastruttura esistente assicurando una buona protezione dell'investimento. Ciò le rende particolarmente adatte per le aziende con i seguenti ambienti:

- ☒ Data center con una crescita incontrollata dei dati di backup e dei fixed content che devono soddisfare requisiti operativi e di disaster recovery sempre più complessi
- ☒ Aziende di grandi dimensioni che prevedono il supporto di uffici remoti per il consolidamento del data center a scopo di protezione dei dati e disaster recovery
- ☒ Infrastrutture di backup e archiviazione che necessitano di una destinazione consolidata per la deduplicazione
- ☒ Infrastrutture basate su sistemi e supporti a nastro con operazioni di backup e ripristino da riprogettare completamente per eliminare problemi di gestione e inutili costi

L'architettura DIA (Data Invulnerability Architecture) di Data Domain assicura l'integrità di tutti i dati di backup grazie agli elevati livelli di protezione, al processo di verifica e alle funzionalità di risoluzione automatica. La protezione dell'integrità dei dati include:

- ☒ Verifica continua della ripristinabilità: controlla che i dati siano corretti e che possano essere ripristinati da ogni livello del sistema durante l'intero ciclo di vita
- ☒ Verifica dell'unicità: protegge i dati da collisioni hash casuali e maligne in modo da garantirne un corretto ripristino
- ☒ Doppia parità del disco di livello RAID 6: garantisce protezione in caso di guasto simultaneo di un massimo di due dischi

L'architettura DIA integra inoltre livelli extra di protezione per l'integrità dei dati in modo da individuare e riparare i guasti per ripristinare i dati di backup senza rischi.

Data Domain Replicator Software trasferisce solo le modifiche univoche deduplicate e compresse su qualsiasi rete IP, il tutto sfruttando, rispetto ai metodi di replica tradizionali, una minore larghezza di banda, un tempo minore e costi inferiori. Se più sistemi Data Domain eseguono la replica sullo stesso sistema di destinazione, la deduplicazione è più efficiente in quanto il sistema di destinazione archivia una sola volta ogni singolo segmento univoco su tutti i flussi di replica in entrata, riducendo ulteriormente la larghezza di banda.

I principali vantaggi della tecnologia di replica Data Domain includono:

- ☒ Vaulting della rete WAN per il disaster recovery: fornisce la protezione dei dati in rete attraverso la replica sicura e automatizzata dei dati di backup su una postazione fuori sede protetta
- ☒ Protezione dei dati in uffici remoti: consente il vaulting dei dati di backup da più uffici remoti a un hub o un data center centrale
- ☒ Replica in cascata: consente la replica dei dati deduplicati da un ufficio remoto a un data center centralizzato e su altri siti, ad esempio quello di disaster recovery, in modo da garantire una migliore ripristinabilità e una maggiore protezione in caso di guasto

- ☒ Consolidamento del nastro: elimina la necessità di deduplicare i dati di backup in ciascun ufficio remoto e mantiene un'infrastruttura basata su nastro notevolmente ridotta solo presso un hub centralizzato con personale IT

---

## **EMC NetWorker**

EMC NetWorker è un'applicazione di backup di classe enterprise che centralizza le operazioni di backup e ripristino. Grazie a NetWorker le aziende possono sfruttare tutte le potenzialità di una piattaforma che supporta un'ampia gamma di opzioni di protezione delle informazioni, inclusi backup su disco, replica, protezione continua dei dati e deduplicazione in ambienti fisici e virtuali. La sua versatilità rende NetWorker il software di backup ideale per i clienti che intendono semplificare la gestione nell'intero ambiente, dai data center di grandi dimensioni agli uffici remoti. L'applicazione alla base di NetWorker fornisce funzionalità di deduplicazione all'origine attraverso l'integrazione con la tecnologia di deduplicazione EMC Avamar ma, nell'ambito del suo funzionamento, è anche compatibile con le soluzioni di deduplicazione sulla destinazione come, ad esempio, i sistemi Data Domain. Le aziende che utilizzano la deduplicazione NetWorker intendono:

- ☒ Ridurre la crescita continua del volume di dati per gli ambienti NetWorker esistenti
- ☒ Sviluppare una nuova strategia di backup su disco volta a migliorare il ripristino che richiede ancora l'utilizzo di un nastro fisico per l'archiviazione o il soddisfacimento dei requisiti a lungo termine
- ☒ Soddisfare una serie di requisiti, alcuni relativi alla deduplicazione all'origine e altri a quella di destinazione
- ☒ Ridurre i costi e la complessità consolidando più strategie di protezione dei dati in un'unica applicazione

In termini di integrazione delle funzionalità di deduplicazione, l'approccio alla deduplicazione offerto da NetWorker è più evoluto rispetto a quello delle tradizionali applicazioni di backup. Il software client di NetWorker per i backup, sia senza deduplicazione che con riconoscimento di quest'ultima, è un singolo agente. Le funzionalità di deduplicazione all'origine sono state completamente integrate, così da ridurre al minimo i requisiti di implementazione e manutenzione. Con la console NetWorker è possibile gestire e monitorare entrambi i tipi di backup – tradizionali e con deduplicazione. Inoltre, le aziende che utilizzano NetWorker e desiderano sfruttare i vantaggi offerti dalla deduplicazione non devono sostenere alcun costo aggiuntivo associato ai client.

A differenza di altre soluzioni, NetWorker non è dotato di unità SKU software incrementali e consente di integrare gratuitamente la tecnologia di deduplicazione. Il motore di deduplicazione più adatto può essere aggiunto con facilità all'ambiente di backup, ad Avamar o alla soluzione back-end Data Domain. Uno dei vantaggi offerti dalla deduplicazione NetWorker consiste nel supporto del nastro fisico: ciò consente agli utenti che utilizzano ancora il nastro di soddisfare i requisiti all'interno della stessa applicazione. È inoltre possibile sfruttare il supporto avanzato delle applicazioni di NetWorker per garantire disaster recovery ottimali e ripristini granulari, oltre che la gestione delle copie

istantanee per il backup fuori sede. NetWorker offre alle aziende potenti funzionalità di deduplicazione che non causano interruzioni dell'ambiente di backup attualmente implementato.

## QUALE APPROCCIO ADOTTARE PER RISOLVERE PROBLEMATICHE SPECIFICHE?

Come descritto in precedenza in questo white paper, ogni approccio o tecnologia di deduplicazione fornisce diversi vantaggi in base all'utilizzo. È quindi importante valutare le singole soluzioni EMC sulla base dell'ambiente aziendale in modo da individuare quale fornisca la massima efficienza. La Tabella 2 aiuta le aziende a capire quale soluzione di backup e ripristino EMC sia la più adatta a seconda delle diverse esigenze.

EMC vanta un ampio portafoglio di prodotti con funzionalità di deduplicazione. Dal momento che questa tecnologia non viene considerata una soluzione standalone, per EMC è importante promuovere la formazione dei clienti nel luogo più appropriato in modo da sfruttarne la capacità sulla base di requisiti ambientali specifici. In combinazione con case study documentati e benchmark sui test per la scalabilità e le prestazioni, la formazione aumenta ulteriormente la familiarità dei clienti nell'applicazione della tecnologia con un determinato prodotto.

**TABELLA 2**

Scelta di una soluzione con deduplicazione EMC

	<b>EMC NetWorker</b>	<b>EMC Data Domain</b>	<b>EMC Avamar</b>
Deduplicazione per il backup	<ul style="list-style-type: none"> <li>• Origine</li> <li>• Deduplicazione in linea</li> </ul>	<ul style="list-style-type: none"> <li>• Destinazione</li> <li>• Deduplicazione in linea</li> </ul>	<ul style="list-style-type: none"> <li>• Origine</li> <li>• Deduplicazione in linea</li> </ul>
Ideale per ambienti:	<ul style="list-style-type: none"> <li>• NetWorker</li> <li>• Con necessità di supportare il nastro fisico</li> <li>• Di grandi dimensioni ed eterogenei</li> </ul>	<ul style="list-style-type: none"> <li>• Elevata velocità nelle operazioni di backup e ripristino</li> <li>• Replica per il backup fuori sede</li> <li>• Supporto dell'attuale ambiente di backup, senza modifiche alle operazioni</li> <li>• Supporto di data center e siti remoti</li> </ul>	<ul style="list-style-type: none"> <li>• Ambienti virtuali</li> <li>• Uffici di filiali remote</li> <li>• Server LAN/NAS</li> <li>• Desktop e laptop</li> </ul>
Opzioni di implementazione	<ul style="list-style-type: none"> <li>• Singolo agente NetWorker</li> <li>• Vedere le opzioni per l'implementazione di Avamar</li> </ul>	<ul style="list-style-type: none"> <li>• Hardware o gateway dell'appliance</li> </ul>	<ul style="list-style-type: none"> <li>• Solo agente: per gli uffici remoti più piccoli</li> <li>• Avamar Data Store: soluzione completa <i>all-in-one</i> (hardware e software)</li> <li>• Server di terze parti: sviluppo di un server Avamar personalizzato</li> <li>• Avamar Virtual Edition: appliance virtuale che sfrutta il server ESX e il disco esistenti</li> </ul>

Fonte: IDC, 2010

## **CONCLUSIONI**

La tecnologia di deduplicazione consente di ottimizzare l'efficienza del backup e ridurre i costi IT. Le aziende possono scegliere tra vari tipi di soluzioni che supportano la deduplicazione, con l'obiettivo di risolvere le problematiche operative e di costi associate alla continua crescita del volume dei dati di backup. IDC considera la deduplicazione come una tecnologia chiave da integrare necessariamente in una serie di soluzioni di storage per soddisfare queste esigenze. Come fornitore, EMC è nella posizione ideale per risolvere i problemi che da tempo affliggono le aziende in quanto offre un'ampia gamma di soluzioni per diversi ambienti e applicazioni in grado di rispondere ai requisiti tecnologici dei clienti nei prossimi cinque anni.

---

## **Copyright**

Pubblicazione esterna di informazioni e dati di IDC – Le informazioni IDC utilizzate in pubblicità, comunicati stampa o materiale promozionale richiedono la preventiva autorizzazione scritta da parte del Vice Presidente o del Country Manager responsabile di IDC. Ogni richiesta in tal senso deve essere accompagnata da una bozza del documento proposto. IDC si riserva il diritto di negare l'approvazione per qualsiasi motivo.

Copyright 2010 IDC. Riproduzione assolutamente vietata senza autorizzazione scritta.