

I vantaggi offerti dalla deduplicazione alle aziende di ogni dimensione

White paper di Acronis[®]



Sommario

Riepilogo	3
Cos'è la deduplicazione?	4
• Deduplicazione a livello di file	
• Deduplicazione a livello di blocco	
• I problemi legati alla protezione	
I vantaggi offerti alle diverse organizzazioni	5
• Vantaggi generali	
• Vantaggi della deduplicazione all'origine	
• Vantaggi della deduplicazione alla destinazione	
Riepilogo	7
Implementazione della deduplicazione Acronis	7
• I vantaggi della deduplicazione Acronis	
• Algoritmo di hash rapido: la chiave per <i>l'ottimizzazione delle prestazioni della deduplicazione all'origine</i>	
• Protezione dei dati deduplicati	
Conclusioni e azioni successive	10

Riepilogo

La crescita dell'archiviazione nelle aziende grandi e piccole è stimata nel 50% - 100% l'anno. Una ricerca condotta da IDC nella seconda metà del 2008 mostra un aumento pari al 3.000% dei dati digitali creati e archiviati su base internazionale negli ultimi tre anni. Inoltre, molte organizzazioni che presentano varie sedi e filiali sono al lavoro per consolidare le proprie risorse dati mediante alcuni sistemi (inclusa la virtualizzazione) che consentano di creare gruppi di risorse con esigenze inferiori in termini di energia, e che possano occupare anche spazi fisici ridotti.

I costi associati all'archiviazione e alla gestione di tutti i dati su disco o su nastro possono essere drasticamente ridotti grazie alla deduplicazione

I vantaggi della deduplicazione dei dati sono stati ampiamente pubblicizzati; essenzialmente, consentono a un'organizzazione di:

- archiviare molti più dati di backup rispetto a un budget stabilito
- aumentare sostanzialmente gli intervalli di acquisto di nuovi dischi
- archiviare su disco in modo efficiente, sfruttando i vantaggi della velocità ed eliminando la necessità del nastro
- ridurre drasticamente gli intervalli di inattività dovuti ai backup.

Se la deduplicazione è una tecnica di riduzione dei dati così efficace, perché non viene utilizzata da ogni organizzazione IT? Fino a poco tempo fa, il costo elevato dei prodotti hardware per la deduplicazione, in genere con marchi proprietari, non consentiva alle aziende, piccole e grandi, di prendere in considerazione tale tecnologia. La stessa preoccupazione a livello di costi imponeva alla relativamente piccola percentuale di organizzazioni che invece potevano acquistarla di riservarla unicamente ai dati dei server, malgrado il fatto che i dati delle workstation rappresentano spesso la metà di tutti i dati posseduti dalle aziende. L'avvento della deduplicazione basata solo su software ha tuttavia sostanzialmente ridotto la soglia di acquisto, rendendo la tecnologia conveniente alle organizzazioni di ogni dimensione e consentendo la deduplicazione anche dei dati delle workstation.

In questo white paper verrà illustrata la tecnologia della deduplicazione, ne verranno esaminati in dettaglio i vantaggi e verrà esposta un'analisi del suo impiego in ambienti Windows® e Linux®.

Cos'è la deduplicazione?

La deduplicazione è concepita per eliminare i dati ridondanti in un sistema di archiviazione e per ridurre la quantità di dati che deve essere archiviata come backup. Può essere applicata a livello di file o di blocco.

In cosa differiscono tali livelli?

La deduplicazione a livello di file

La deduplicazione a livello di file individua i file che sono esattamente simili e ne archivia una sola copia, posizionando dei 'puntatori' al posto delle copie. Benché questa tecnica sia comunque più efficiente del non applicare alcuna forma di deduplicazione, è bene tenere presente che ad ogni piccola modifica apportata a un file corrisponderà una copia aggiuntiva archiviata.

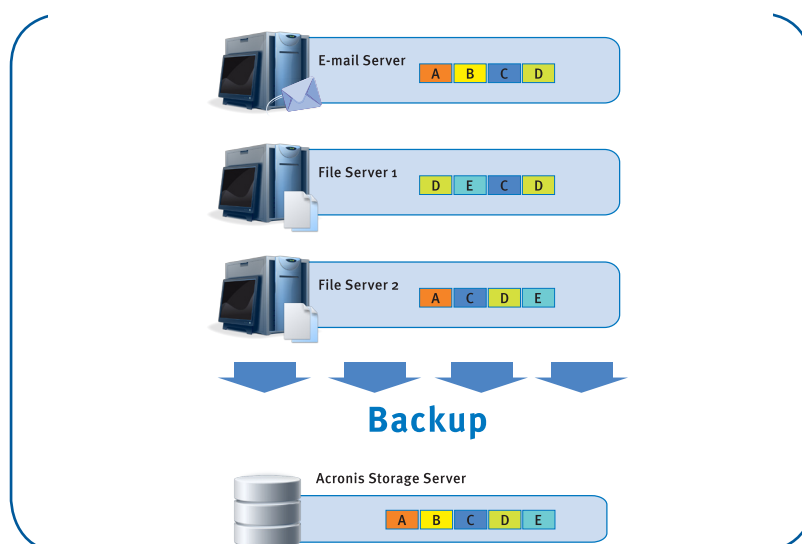
Deduplicazione a livello di blocco

Deduplicazione a livello di blocco. La deduplicazione a livello di blocco sembra promettere un'efficienza complessiva maggiore dell'archiviazione. Funziona mediante la ricerca di istanze di informazioni ridondanti tramite l'individuazione di blocchi di dati con dimensione pari a 4 KB e più, dei quali viene archiviata una sola copia, indipendentemente dal numero di copie che vi sono. Le copie vengono sostituite da puntatori che fanno riferimento al blocco di dati originale in una modalità praticamente invisibile all'utente, che continuerà a utilizzare un file come se tutti i blocchi di dati che contiene fossero originali.

La deduplicazione riduce il volume dei dati archiviati fino al 90%

Per comprendere la potenza di questa tecnologia, si consideri l'effetto sul sistema di backup dell'invio mediante email a dieci colleghi dell'azienda di una presentazione di Microsoft® Powerpoint® piena di immagini, che occupa 9 MB. Quando si preme il pulsante 'Invia', vengono clonate 10 copie di quel file da 9 MB. Quando viene eseguito il backup dei dati di ogni destinatario mediante tecniche tradizionali, ogni istanza della presentazione viene memorizzata e archiviata. Ciò significa che in poco tempo un file da 9 MB occuperà in realtà 99 MB di spazio di storage. Moltiplicate questo dato per le centinaia di altre istanze di dati clonati che si hanno ogni giorno e inizierete a capire perché i requisiti di archiviazione su disco, e i costi a questi associati, mostrano ormai una crescita vertiginosa.

La deduplicazione è una tecnica comprovata per ridurre i costi di avvio delle strutture di storage e per risparmiare al contempo sulla larghezza di banda della rete. Rende possibile sia aumentare la capacità di storage dei dati per ogni singola unità di archiviazione (allungando ad esempio i tempi di acquisto di ulteriori unità disco), sia di conservare i dati online per periodi più lunghi.



Gli utenti possono iniziare ad investire di più, non solo in capacità concreta, ma in strumenti che ottimizzano l'utilizzo dello storage [ad esempio, thin provisioning, deduplicazione dei dati e virtualizzazione dello storage].
Natalya Yezhkova, Research Manager, Storage Systems, IDC. Gennaio 2009

L'intera procedura (sinistra) può essere eseguita come:

- *una funzione all'origine*, dove i duplicati vengono eliminati prima di essere scritti sul disco di destinazione, oppure
- *una funzione alla destinazione*, che identifica i dati duplicati già scritti su disco e li rimuove.

Perché è importante?

Per molte aziende la deduplicazione significa la riduzione del volume dei dati, in modo che, sostanzialmente, tutti i backup possano essere conservati su disco, ovviando alla necessità dei nastri e dello storage offline, ad eccezione forse degli archivi a lungo termine. Con questa transizione, gli amministratori possono effettuare backup e ripristini più rapidi, coerenti con le soluzioni di protezione dei dati basate su disco. La tecnologia semplifica inoltre la conformità ai requisiti governativi e finanziari, che impongono l'archiviazione di tutte le copie di documentazione prodotta nel corso di più anni.

Aspetti relativi all'integrità dei dati

Se la deduplicazione consente di salvare grandi quantità di spazio, la sua concentrazione ne può rendere difficoltosa l'adeguata archiviazione. Se un blocco di dati individuato su diverse origini (come nel precedente esempio della presentazione di Powerpoint) viene deduplicato e poi va perso, risulteranno danneggiati tutti i backup ad esso associati, poiché il dato del backup di origine non esiste più. Questo vale anche per i backup di sistemi completi. Nel caso in cui si verifichi un'emergenza, un singolo blocco dati danneggiato che però corrisponde a una parte vitale di un sistema operativo Windows renderà inutilizzabili tutti i backup destinati al ripristino del sistema stesso. Considerare l'utilizzo di un array RAID per archiviare i dati della deduplicazione allo scopo di fornire un ulteriore livello di protezione.

Quali sono i vantaggi offerti alle diverse organizzazioni?

Vantaggi generali

La deduplicazione promette alle organizzazioni di archiviare un numero molto più elevato di dati per unità di storage di quanto non sia stato possibile finora. In alternativa, per una spesa equivalente, si può scegliere di conservare i dati online per periodi di tempo più lunghi. In ogni caso, ciò si traduce in diversi vantaggi aziendali:

- **riduzione del costo complessivo dello storage** (perché il volume dei dati archiviati è inferiore)
- **aumento effettivo della larghezza di banda di rete** (non è necessario trasmettere copie sulla rete se la deduplicazione avviene all'origine)
- **un contesto più rispettoso dell'ambiente** (meno elettricità, meno metri cubi di spazio richiesto per ospitare i dati, sia nelle sedi principali che in quelle remote)
- **ripristini più rapidi**, a garanzia della continuità dei processi aziendali
- **possibilità di garantire la conformità ai requisiti di archiviazione dei dati aziendali e legali**, senza che le dimensioni dello storage aumentino a dismisura
- **rapido ritorno sugli investimenti** (perché si acquista e si esegue la manutenzione di una quantità inferiore di supporti di storage)
- **riduzione degli intervalli di backup** (il backup dei puntatori ai dati, piuttosto che delle copie dei dati, richiede solo una frazione ridotta dello spazio necessario).

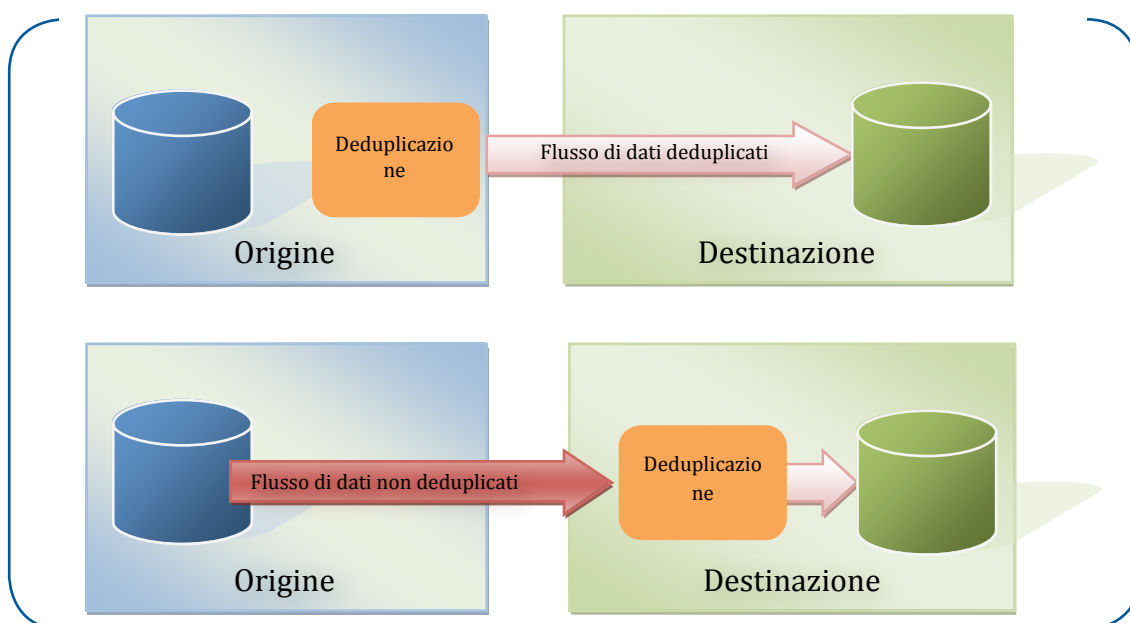
Vantaggi specifici della deduplicazione all'origine

La deduplicazione all'origine (o sul lato server) mostrata nella parte superiore dell'immagine seguente, consente di:

- ridurre la quantità di dati trasferiti in una rete in una posizione di storage di destinazione da 10 a 20 volte.
- elimina i potenziali colli di bottiglia dovuti alla trasmissione, in particolare negli scenari in cui le reti esistenti già operano alla massima capacità, oppure quando si eseguono backup di posizioni remote su linee di comunicazione con larghezza di banda limitata.
- è efficace per tutti i tipi di dati archiviati, siano essi associati alle applicazioni o meno.
- meno complessità di implementazione, poiché non richiede hardware aggiuntivo o client sul lato di destinazione.

Svantaggio principale.

L'esecuzione dei backup può richiedere più tempo e più cicli di CPU durante le procedure di deduplicazione dei dati, aggiungendo potenziali problematiche legate alle prestazioni delle macchine di produzione. Tuttavia, come verrà esaminato più avanti nel documento, una nuova tecnologia, denominata *ottimizzazione delle prestazioni della deduplicazione all'origine* può eliminare la maggior parte degli effetti collaterali della performance di questo tipo



*Errore. L'origine riferimento non è stata trovata.
Deduplicazione all'origine e alla destinazione a confronto*

Vantaggi specifici della deduplicazione alla destinazione (parte inferiore dell'immagine).

La deduplicazione alla destinazione avviene, dopo l'esecuzione del backup dell'origine, nella posizione di storage di destinazione, che in genere è un nodo di archiviazione connesso.

Vantaggio principale.

Il primo backup all'origine può essere completato più rapidamente spostando le attività di deduplicazione ad elevata richiesta di CPU dalla macchina di origine, abbreviando così la finestra di backup. Questa variante viene spesso preferita nelle situazioni in cui gli amministratori operano con client e origini di dati non predisposti per la deduplicazione, o quando il carico di elaborazione dei dati associati alla deduplicazione all'origine allunga i tempi di backup oltre i limiti di tempo previsti dagli amministratori.

Svantaggio principale.

È necessario trasmettere mediante la rete tutte le copie che esistevano precedentemente alla deduplicazione, causando potenziali colli di bottiglia sulla larghezza di banda. La scelta tra le due tipologie di deduplicazione dipenderà da quale limitazione - *carico di lavoro della CPU del client o considerazioni relative alla larghezza di banda* - risultano essere prioritarie per l'organizzazione.

Riepilogo

La deduplicazione è stata per lungo tempo uno strumento esclusivo delle grandi aziende, caratterizzata da costi proibitivi, difficoltose curve di apprendimento e backup basati sulla deduplicazione solo a livello di file, una funzionalità che limitava l'utilizzo dei dati deduplicati per il ripristino di una macchina danneggiata. Fino a oggi, questa tecnologia è stata così costosa da poter essere implementata soltanto dalle più grandi organizzazioni. Inoltre, poteva essere applicata solo a supporto dei server, malgrado il fatto che, nella maggior parte delle infrastrutture IT, la maggior parte dei dati siano contenuti presso le workstation.

La maggior parte dei prodotti di deduplicazione sono stati ideati e venduti come soluzioni software/hardware combinate. In quasi tutti i casi era difficile giustificare i costi elevati anche del solo hardware. Per illustrare questo ultimo punto, si consideri il fatto che nel marzo del 2009 uno dei fornitori più noti ha ridotto di oltre un terzo il costo di uno dei suoi dispositivi high-end per la deduplicazione dei dati. Eppure, a un prezzo documentato di 130.000 dollari per 12 TB di capacità di storage, si tratta di una proposta ancora eccessivamente costosa. Ostacoli come questi hanno limitato le promesse della deduplicazione solo alle organizzazioni più grandi.

Oggi questi limiti sono finalmente superati, e la deduplicazione può essere specificata più ampiamente:

- non solo dalle grandi enterprise, ma anche da organizzazioni più piccole che devono superare ostacoli significativi rispetto allo storage dei dati.
- non solo per i server, ma anche per le workstation.

Presentazione di Acronis® Backup & Recovery™ 10 Deduplication

La funzione di deduplicazione integrata in alcune soluzioni Acronis ha diversi vantaggi che la distinguono dalle offerte di molti altri produttori:

- **L'utente può scegliere se applicarla all'origine o alla destinazione.**
Molte organizzazioni devono poter essere in grado di implementare entrambe le possibilità in reparti diversi.
- **Backup a livello di file e di blocchi.**
I backup a livello dei soli file limitano enormemente il risparmio potenziale possibile con la deduplicazione.
- **La deduplicazione rapida all'origine.**
Elimina spesso la necessità della deduplicazione alla destinazione. Acronis ha compiuto un grande passo avanti nel ridurre il carico di CPU nella deduplicazione all'origine, eliminando la necessità per molte organizzazioni di spostare i processi presso server o dispositivi di deduplicazione alla destinazione.
- **Conveniente.**
Molte organizzazioni possono giustificare i costi della deduplicazione non solo dal punto di vista dei server, ma anche da quello delle workstation.
- **Integrata con i prodotti di backup e ripristino di Acronis.**
Funziona senza soluzione di continuità con il software Acronis Backup & Recovery 10, così che i dati deduplicati risulteranno protetti al pari di quelli non deduplicati.

L'opzione Acronis Backup & Recovery 10 Deduplication viene fornita come funzionalità completamente integrata nella famiglia di prodotti della nuova soluzione software Acronis Backup & Recovery 10. Si tratta di una soluzione solo software che può essere acquistata con le seguenti edizioni avanzate di Acronis Backup & Recovery 10:

- Advanced Server Edition
- Advanced Workstation Edition
- SBS (Small Business Server) Edition
- Virtual Edition

I vantaggi della deduplicazione Acronis

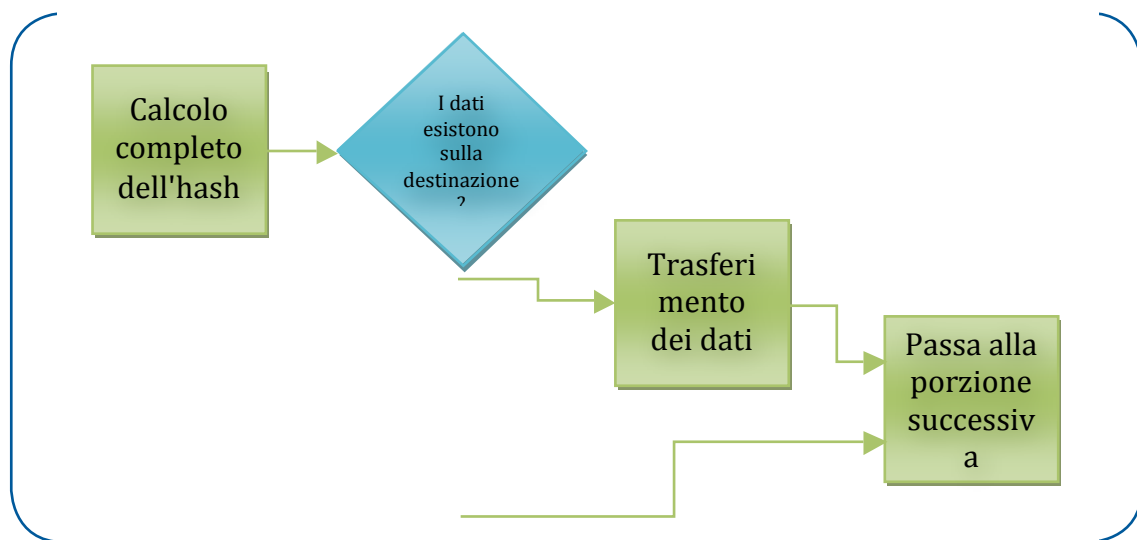
A differenza di molte altre soluzioni, Acronis supporta la deduplicazione all'origine e alla destinazione. Ma si distingue anche per altre caratteristiche:

- **Backup basato sull'immagine.**
La funzione di deduplicazione dei dati offre risparmi notevoli su costi di archiviazione e larghezza di banda di rete.
- **Soluzione completamente integrata con il software per il ripristino d'emergenza.**
Ripristino di file e sistemi – deduplicati e non - in pochi minuti, anziché ore o giorni. Elimina l'archiviazione nelle immagini di più copie di grandi origini dati, ad esempio i sistemi operativi multi-gigabyte.
- **Più tipologie di backup.**
La deduplicazione può essere applicata ai backup completi, incrementali o differenziali.
- **Convenienza economica**
La tecnologia di deduplicazione software di Acronis è conveniente.
- **Utilizzo dell'hardware di archiviazione esistente.**
L'opzione di deduplicazione di Acronis elimina l'esigenza di hardware proprietario costoso.
- **Installazione rapida.**
In genere è funzionante in circa un'ora, e non servono i diversi giorni necessari per l'installazione dei sistemi hardware/software.
- **Facile da utilizzare.**
La stessa semplicità di utilizzo e i limitati requisiti di formazione che distinguono i prodotti Acronis caratterizzano anche l'opzione di deduplicazione.

L'efficienza di archiviazione senza precedenti è la realtà, soprattutto se combinata con i potenti algoritmi di compressione Acronis (e con le altre funzioni orientate all'efficienza) disponibili in Acronis Backup & Recovery 10. Se utilizzata congiuntamente alla potente utilità di compressione dei dati Acronis, gli amministratori IT possono ridurre ulteriormente i costi complessivi di archiviazione dei dati, *dopo la deduplicazione*, in media del 50% - 60%, a seconda del tipo di file, generando ulteriori risparmi sostanziali di storage su disco. Comprimendo i dati durante la procedura di repack, si alleggerisce il carico dell'elaborazione dai server della linea di produzione. Vengono compressi sia i file .tib di Acronis che i blocchi di archiviazione dei dati deduplicati.

Algoritmo di hash rapido: la chiave per l'ottimizzazione della deduplicazione all'origine

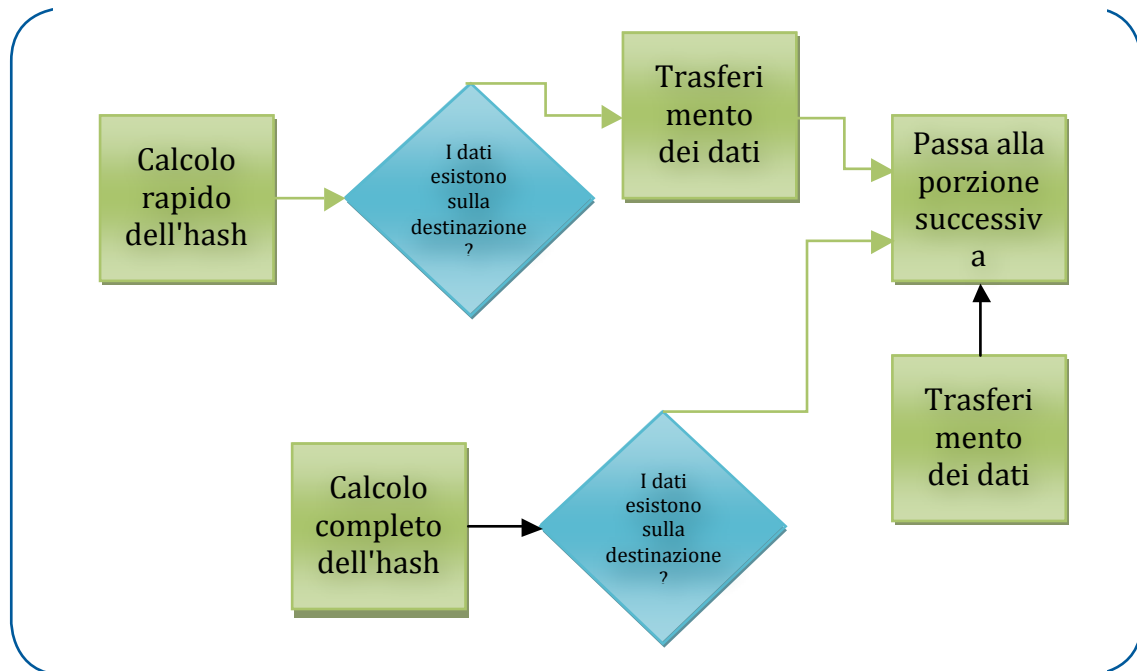
Acronis offre un approccio più efficiente alla deduplicazione all'origine. Per illustrare ciò che abbiamo realizzato, si osservi un algoritmo di deduplicazione all'origine standard (di seguito). In questa situazione, il client calcola innanzitutto il checksum dei dati di cui deve essere eseguito il backup. Il risultato viene definito hash. Tale hash viene inviato alla destinazione, che risponde dichiarando "Non ho i dati" oppure "Ho i dati." Nel primo caso, il client invia alla destinazione i dati correnti prima di procedere alla porzione successiva. Nel secondo caso, non è necessaria alcuna ulteriore azione da parte del software client, ed è possibile elaborare la porzione successiva dei dati, come mostrato di seguito.



*Errore. L'origine riferimento non è stata trovata.
Deduplicazione all'origine standard*

Purtroppo, la deduplicazione standard all'origine implica un carico di lavoro significativo, poiché calcola *sempre* l'hash, a prescindere dal fatto che sulla destinazione siano presenti o meno i dati. Ciò è necessario, perché la destinazione non è in grado di dichiarare se i dati sono già disponibili prima che l'hash venga fornito dall'origine. Nei sistemi caratterizzati da carichi di lavoro consistenti, questo tipo di deduplicazione può creare rallentamenti di sistema che spesso portano i responsabili IT a non utilizzarla affatto.

Acronis offre un approccio molto meno esigente in termini di utilizzo della CPU, che rende la deduplicazione all'origine un'opzione fattibile per la maggior parte delle aziende. Viene definita ottimizzazione delle prestazioni della deduplicazione all'origine. Si tratta di un potente algoritmo che elimina la maggior parte dei calcoli dell'hash per i dati che devono ancora essere scritti sulla destinazione.



Errore. L'origine riferimento non è stata trovata.

Ottimizzazione delle prestazioni della deduplicazione all'origine di Acronis

In questo approccio, Acronis crea prima un hash rapido selezionando una piccola quantità di dati che ha maggiori probabilità di venire cambiata quando il dato viene modificato. L'hash è molto rapido e risponde dichiarando "Non ho i dati" oppure "Potrei avere i dati." Nel primo caso, i dati correnti vengono inviati al client. Nel secondo caso, viene calcolato l'hash completo, garantendo così che la destinazione risponda in modo affidabile.

Per maggiore sicurezza, i dati deduplicati vengono crittografati. Durante la creazione dei depositi è possibile specificare una password per la crittografia del deposito – protetta negli storage protetti di Windows. I dati crittografati saranno accessibili solo mediante tale password, e qualsiasi tentativo di recuperarli dal deposito di archiviazione dei dati deduplicati sarà inutile.

Conclusioni e azioni successive

Benché Acronis non sia la prima azienda ad offrire la deduplicazione, la nostra tecnologia basata sull'immagine, con i suoi backup rapidi e i ripristini praticamente immediati, porta la deduplicazione a un livello totalmente nuovo, applicabile ai dati di backup a livello di sistema e di file, ai server e alle workstation. Acronis rende la deduplicazione più accessibile, sia dal punto di vista finanziario che da quello della facilità d'uso, e a un maggior numero di utenti. Quando utilizzata insieme ad Acronis Backup & Restore 10, ridefinisce anche la protezione dei dati. Di seguito alcune indicazioni per avvicinare la vostra organizzazione alla deduplicazione:

- 1:** Provate il nostro calcolatore di deduplicazione sul sito Web di Acronis. Potrete rapidamente stabilire quanto vi è possibile risparmiare utilizzando il software Acronis Backup & Recovery 10 Deduplication.
- 2:** Provate la soluzione scaricando la versione di prova. Per poterla utilizzare sarà necessario scaricare anche Acronis Backup & Recovery 10.
- 3:** Troverete ulteriori informazioni sul nostro sito web www.acronis.com, oppure contattandoci ai numeri di telefono elencati alla fine di questo documento.



ACRONIS
VIA VALLA 16
20141 MILANO, ITALIA

Per ulteriori informazioni, visitate il sito <http://www.acronis.it>

Per clienti aziendali e richieste sul canale:

Tel: +39 02 84742-352

Fax: +39 02 84742-353

Copyright © 2000-2009 Acronis, Inc. Tutti i diritti riservati. "Acronis", "Acronis Compute with Confidence", "Acronis Backup & Recovery" e il logo Acronis sono marchi Acronis, Inc. Windows è un marchio registrato di Microsoft Corporation. Linux è un marchio registrato di Linus Torvalds. Gli altri nomi menzionati possono essere marchi o marchi registrati dei rispettivi titolari. Soggetto a modifiche tecniche. Le immagini potrebbero non corrispondere al prodotto reale. Si declina qualsiasi responsabilità per possibili errori. 2009-08