

BACKUP



1. Perché il BACKUP



Il funzionamento delle attività aziendali si basa sulla corretta gestione dei ***dati critici***.

Si richiede:

- un aumento dei requisiti di archiviazione dati
- soluzioni per la protezione dei dati
- soluzioni per la gestione dei dati dislocati in diverse aree geografiche
- disponibilità ininterrotta dei dati

0. BACKUP: gli obiettivi



- Assicurare la disponibilità e continuità dei servizi
- Gestire la crescita dei volumi dati
- Gestire l'infrastruttura di archiviazione per migliorare la qualità del servizio riducendo la complessità e i costi
- Integrare le applicazioni con i requisiti di archiviazione e gestione dati
- Eseguire il backup dei dati entro intervalli di tempo brevi
- Supportare i sistemi di IT esistenti (legacy) che non consentono l'esecuzione delle tecnologie più recenti
- Determinare il valore dei dati in modo da applicare le strategie più appropriate

1. La perdita dei dati può essere causata da:



- Errore di un sottosistema del disco rigido
- Interruzione dell'alimentazione con conseguente danneggiamento dei dati
- Danneggiamento fisico dei supporti
- Errore del software di sistema
- Eliminazione oppure modifica accidentale o non autorizzata dei dati
- Virus
- Disastri naturali
- Furto o sabotaggio

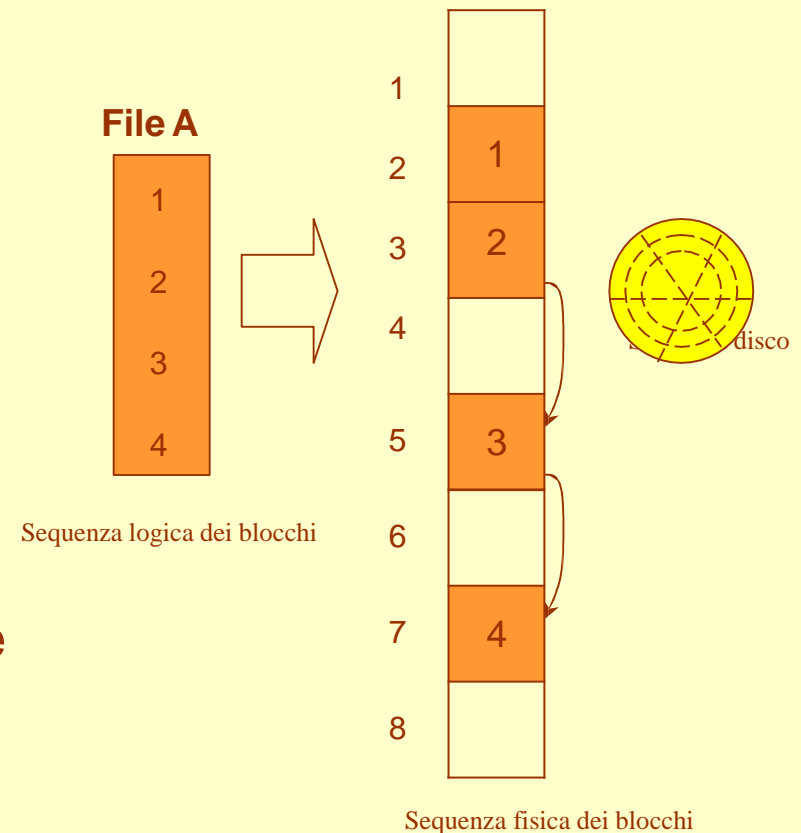
2. Organizzazione fisica dell'informazione su disco

I dati su disco rigido sono organizzati in blocchi di piccole dimensioni.

Un file è costituito da più blocchi fisici *non necessariamente sequenziali*.

In alcuni sistemi operativi l'associazione tra blocchi fisici e filename è realizzata tramite una lista concatenata:

- Nella FAT (o altra struttura NFTS etc.) è memorizzato l'indirizzo del primo cluster
- In coda ad ogni cluster è riportato l'indirizzo del successivo



2. Organizzazione fisica dell'informazione su disco

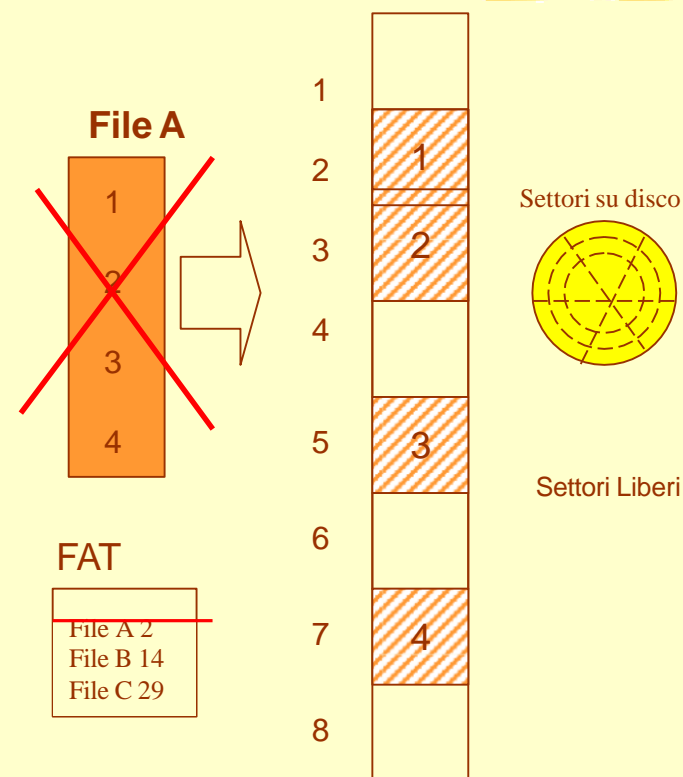
- Cancellazione di un file è solo logica
- Dati restano su disco anche se si effettua vuota cestino

Punto di controllo sicurezza:

- vecchi computer
- vecchi supporti
- vecchi cd

Buona pratica: smaltimento dei dati

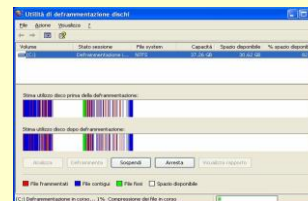
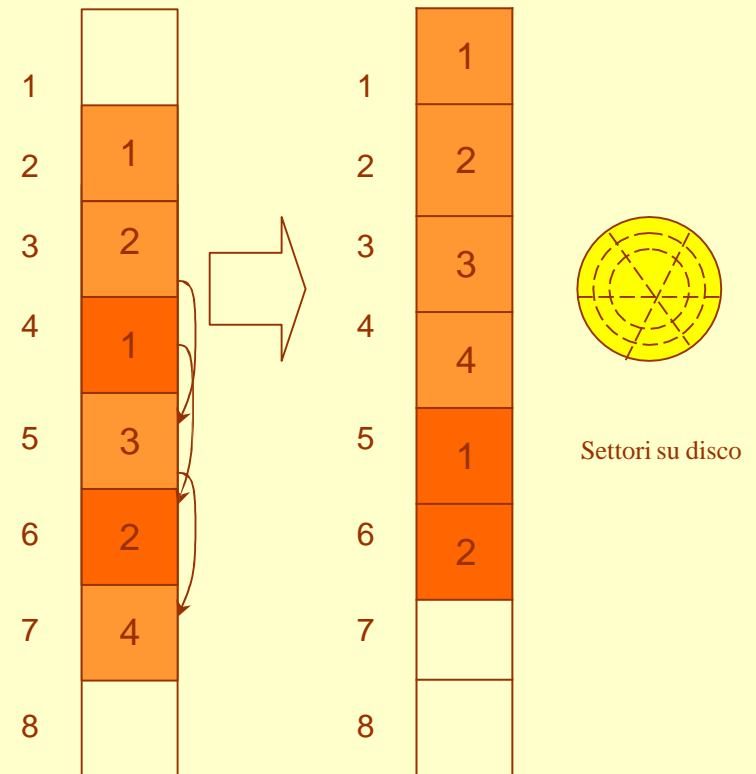
- non gettare/riutilizzare i supporti ma prima effettuare cancellazione fisica con appositi strumenti
- frantumatori



2. Operazioni su disco: defrag

Defrag: riorganizza i settori del disco in cui sono memorizzati i dati, eliminando gli intervalli vuoti che si creano a causa delle ripetute operazioni di lettura/scrittura/cancellazione dei file

Dopo deframmentazione lettura da disco più veloce in quanto i settori logicamente consecutivi di un file risulteranno anche fisicamente consecutivi.

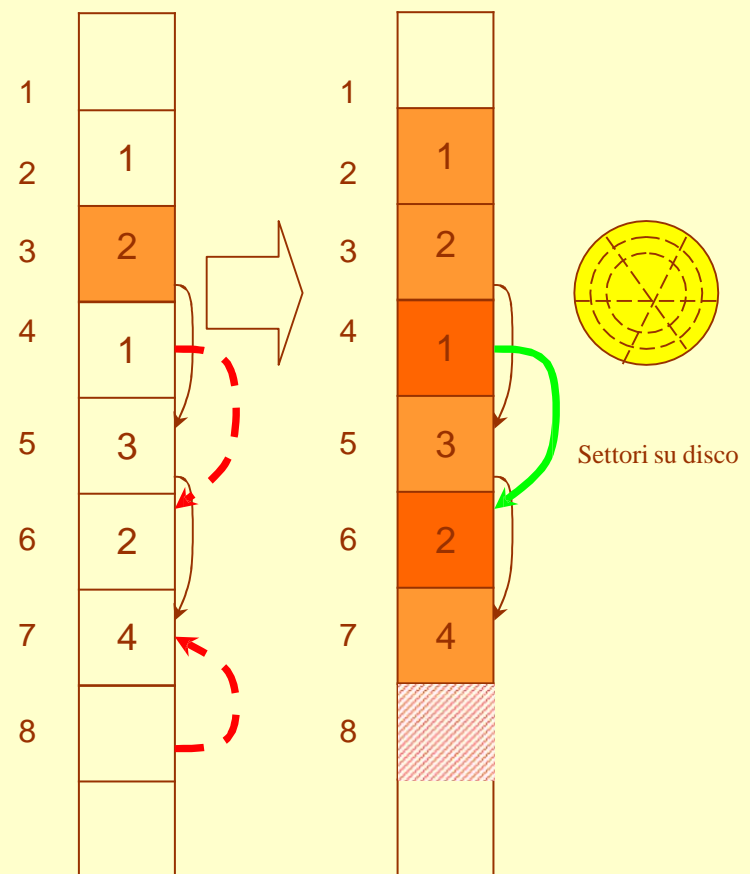


Riorganizzazione e Compattamento

2. Operazioni su disco: scandisk

Scandisk: scandisce il file system del disco alla ricerca di eventuali file o settori danneggiati a causa di errate operazioni di scrittura

Utile per evitare che un errore nella struttura del file system si propaghi all'interno del sistema.




3. Configurazione dei servizi di backup e ripristino



Per una progettazione di una soluzione di backup efficiente si deve:

- pianificare una **strategia di backup** appropriata per l'ambiente specifico
- decidere la **modalità di backup** più appropriata per l'ambiente specifico
- decidere il **tipo di backup** più appropriato per l'ambiente specifico
- decidere la **tipologia di backup** più appropriata per l'ambiente specifico

3. Strategie di backup: Esclusione dei backup non necessari




Maggiore è il numero dei file sottoposti a backup, maggiore sarà il tempo necessario per l'esecuzione del processo e per il ripristino dei file.

I backup di notevoli quantità di dati eseguiti regolarmente influiscono negativamente sulle prestazioni di rete (a meno che si usi una rete di backup dedicata).

L'obiettivo è quello di ripristinare correttamente l'ambiente dopo un'interruzione quindi:

- i dati da ripristinare devono essere facilmente individuabili
- il ripristino deve essere quanto più rapido possibile

3. Strategie di backup: scelta dell'orario più appropriato per il backup




In ogni tipo di ambiente esistono diverse possibilità per eseguire un backup efficiente.

La scelta di un orario appropriato per il backup è quella di creare un **disagio minimo per gli utenti**.

- Pianificare il backup per evitare gli orari di utilizzo massimo (il backup in un ambiente di e-commerce o in una rete LAN aziendale)
- Non effettuare il backup di dati non necessari
- Effettuare operazioni regolari di ripristino di prova su una rete di test per controllare la correttezza dei backup

3. Strategie di backup: scelta del supporto di archiviazione appropriato



Per scegliere un supporto di archiviazione adeguato si deve considerare:

|

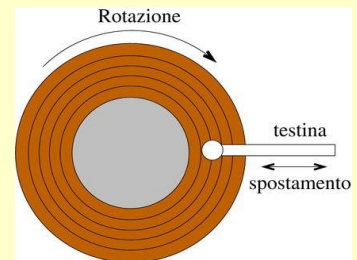
- La quantità di dati da sottoporre a backup
- tipo di dati da sottoporre a backup
- intervallo di backup
- ambiente
- distanza tra i sistemi sottoposti a backup e la periferica di archiviazione, valutazione rischi
- interruzione
- budget

3. Strategie di backup: Scelta del supporto di archiviazione appropriato – Memoria a stato solido

- Memorizzare di un volume NON elevato di dati .
- Anche se esterni, possono essere collegati al computer tramite connessioni ad alta velocità
- Velocità inferiore a dischi rigidi
- Supporto più costoso per l'archiviazione iniziale.

3. Strategie di backup: Scelta del supporto di archiviazione appropriato - Disco Rigido

- Memorizzare di un volume elevato di dati .
- Anche se esterni, possono essere collegati al computer tramite connessioni ad alta velocità
- Sono veloci e consentono di eseguire le operazioni di backup rapidamente.
- Sono facili da configurare e da gestire.
- Supporto meno costoso / Terabyte
- E' anche possibile utilizzare un secondo disco rigido interno o utilizzare una tecnologia RAID; in questo caso però la copia di backup resta all'interno del computer



3. Deterioramento dei supporti: problemi dei vecchi supporti (legacy)

I **nastri** soffrono l'umidità.

Per i **CD** le principali cause di deterioramento sono:

- Il calore che è in grado di renderli illeggibili;
- sono insensibili ai campi magnetici ad esempio prodotti dal monitor o dalle casse (campi che invece danneggiano, magneti ottici, e HD magnetici)
- i **CDR** e **CDrw** sono sensibilissimi all'esposizione diretta ai raggi del sole, che in pochissimo tempo li può "cancellare"
- i **CD, CDR e CD-RW** hanno la capacità di correggere gli errori dovuti a graffi, polvere o altro entro ovviamente certi limiti.



**EVOLUZIONE CONTINUA: PROGRAMMARE PERIODICAMENTE
BACKUP di supporti legacy con le nuove tecnologie**

4. Modalità di backup



La modalità di backup dipende dai dati da sottoporre a backup e determina l'esecuzione del processo:

Backup in linea: i backup vengono eseguiti mentre i dati sono ancora accessibili agli utenti.

Backup non in linea: i backup vengono eseguiti sui dati già resi inaccessibili agli utenti

4. Modalità di backup: Backup in linea



Vengono eseguiti quando il sistema è in funzione, interruzione minima.

Vantaggi:

- nessuna interruzione dei servizi (applicazioni e dati sempre disponibili)
- il backup può essere effettuato in qualsiasi orario
- backup completo o parziale

Svantaggi

- prestazioni del server (riduzione delle prestazioni)
- file aperti (a seconda delle applicazioni attive, è possibile che alcuni file aperti non vengono sottoposti a backup causa **lock in lettura**)

4.Modalità di backup: Backup non in linea



Vengono eseguiti quando il sistema o i servizi non sono in linea.

Vantaggi

- Backup completo o parziale
- Prestazioni
- Backup di tutti i file

Svantaggi

- Interruzione dei servizi (i dati non sono accessibili agli utenti mentre è in corso il processo di backup, per evitare errori in **transazioni**)

4. Tipi di backup



Per i backup in linea e non in linea è possibile utilizzare vari tipi di backup:

- ***Backup completi***
- ***Backup incrementali***
- ***Backup differenziali***

La scelta dipende da:

- SLA (Service Level Agreement) concordato
- intervallo dei backup
- tempo di ripristino.

4. Tipi di backup: Backup completi

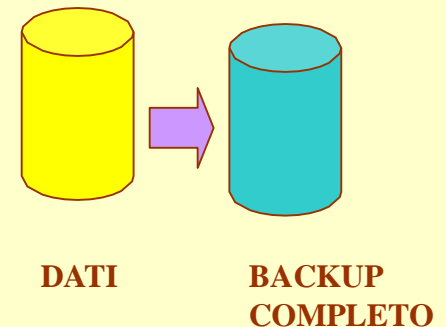
Si acquisiscono tutti i dati, inclusi i file di tutte le unità disco rigido. E' possibile utilizzare una **copia di backup completo** aggiornata per ripristinare completamente un server

Vantaggi

- Copia completa dei dati
- Accesso rapido ai dati di backup

Svantaggi

- Dati ridondanti
- Tempo per eseguire backup



4. Tipi di backup: Backup incrementali

Si acquisiscono tutti i dati modificati dopo un backup completo o incrementale più recente. Per ripristinare un server è necessario:

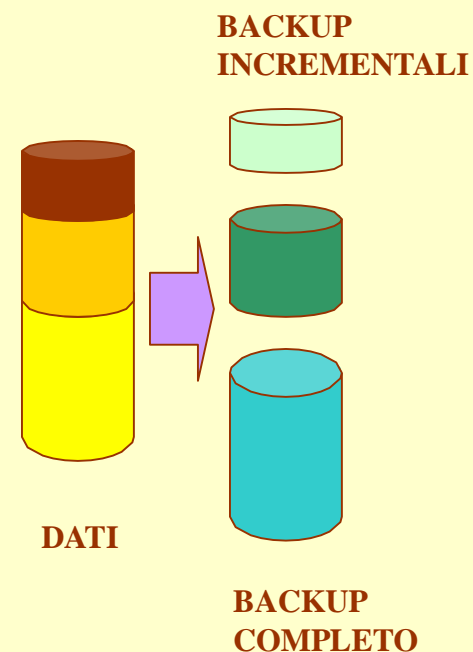
- il supporto del backup completo e
- tutte le serie successive di backup incrementali.

Vantaggi

- Utilizzo efficiente del tempo
- Utilizzo efficiente dei supporti di backup (meno spazio per gli incrementi)

Svantaggi

- Complessità dei ripristini completi
- Ripristini parziali più lunghi



4. Tipi di backup: Backup differenziali

Un backup differenziale acquisisce i dati modificati dopo l'ultimo backup completo. Per ripristinare un server è necessario

- Il supporto di backup completo e
- Il supporto del backup differenziale più recente.

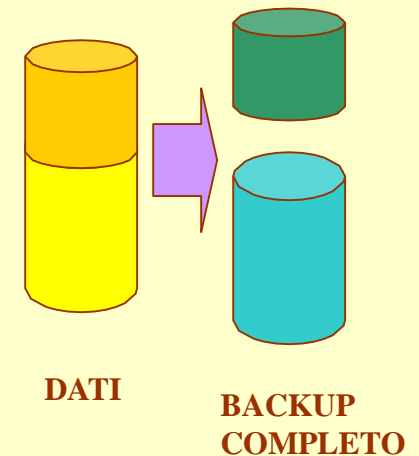
Vantaggi

- Ripristino rapido (minor numero di supporti, per un ripristino completo servono al più due supporti)

Svantaggi

- Backup più lunghi e di maggior dimensioni
- Aumento progressivo dimensione backup differenziale
- Aumento del tempo di backup

BACKUP DIFFERENZIALE



5. Tipologie di backup



E' possibile classificare le tipologie di backup e ripristino in base alla tecnologia di archiviazione (DAS, NAS o SAN) da sottoporre a backup.

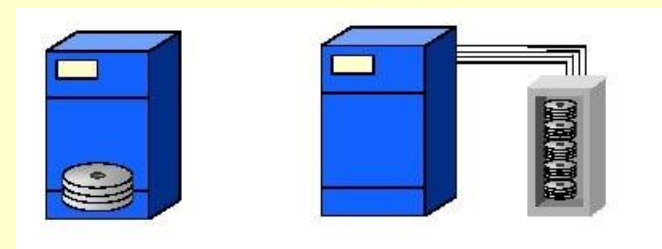
Le tipologie relative a ogni tipo di archiviazione sono rispettivamente:

- Il backup del server locale **DAS** (Direct Attached Storage)
- Il backup **NAS** (Network Attached Storage) collegati alla rete LAN
- I sistemi **SAN** (System Area Network)

5. Tipologie di backup: Backup e ripristino del server locale

Ogni server è connesso alla relativa periferica di backup, tramite un bus SCSI:

Directly Attached Storage (DAS)



Vantaggi

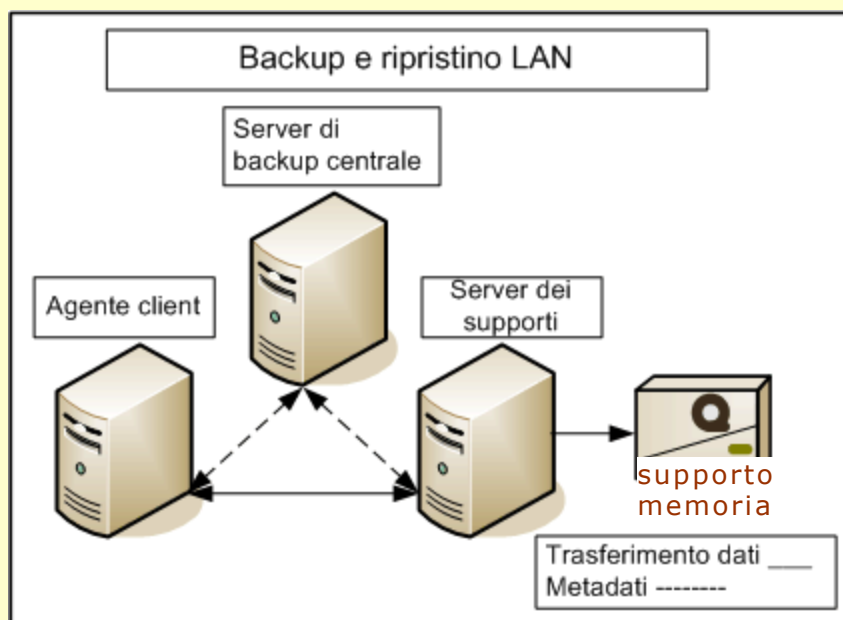
- Nessun consumo delle risorse di rete
- Backup e ripristino più rapidi

Svantaggi

- Funzionalità limitate per la gestione centralizzata e la scalabilità
- Costi più elevati per il software di backup e l'hardware delle periferiche dedicate

5. Tipologie di backup: Backup e ripristino LAN

Il software per questo tipo di backup utilizza un'architettura a più livelli in cui alcuni server di backup avviano i processi e raccolgono i metadati ai dati sottoposti a backup (dati di controllo) mentre altri server eseguono l'operazione effettiva di gestione dei dati trasmessi alle periferiche che gestiscono i supporti.



- server di backup centrale (ospita il modulo di backup che controlla l'ambiente)
- server di supporto (gestisce o spostamento dei dati e le risorse dei supporti)
- agente client (è un agente specifico dell'applicazione)

5. Tipologie di backup: Backup e ripristino LAN

Vantaggi

- Elevata capacità di memorizzazione
- Applicazioni di backup eseguite su server di backup dedicati
- Agenti client che trasferiscono i dati attraverso la rete LAN ad un server di backup
- Livello più elevato di scalabilità e condivisione di singole periferiche per i supporti di memoria
- Utilizzo di RAID per la protezione

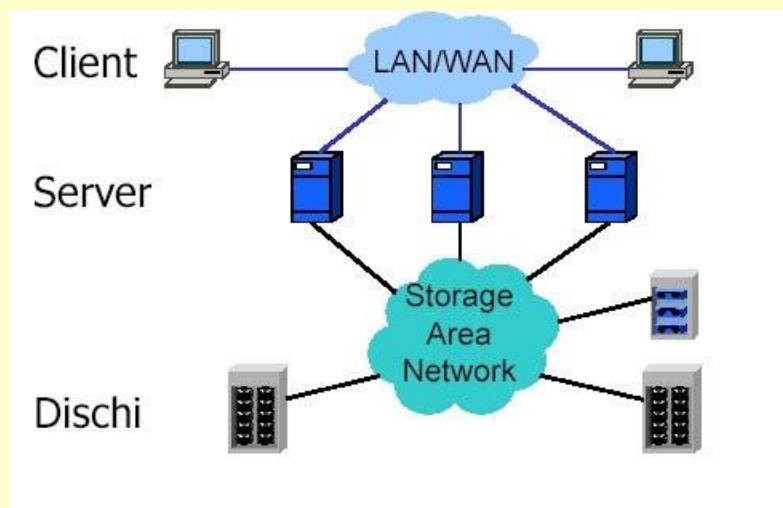
Svantaggi

- Possibile riduzione delle prestazioni del server e della rete
- Pianificazione obbligatoria del processo di backup e ripristino

5. Tipologie di backup: Backup e ripristino SAN

Le reti SAN (System Area Network) consentono la condivisione delle risorse (LAN) ma sono appositamente progettate in modo che i server condividano le periferiche di archiviazione, ad esempio array di dischi.

Spostamento dell'operazione di copia di backup effettiva dall'host di produzione ad un sistema host secondario



Vantaggi

- Carico del server ridotto
- Carico della rete LAN ridotto
- Soluzioni di archiviazione ottimizzate
- Rapidi processi di ripristino dei dati

Svantaggi

- Costi (è necessaria una rete SAN)
- Compatibilità delle periferiche

5. RAID (Redundant Array Independent Disks)



La tecnologia RAID, insieme di dischi ridondanti e indipendenti, ha lo scopo di immagazzinare gli stessi dati in posti differenti su dischi multipli per migliorare le prestazioni e/o tolleranza ai guasti (fault-tolerance).

- **Migliori prestazioni:** dividendo in parallelo il carico di lavoro su più dischi fisici
- **Tolleranza ai guasti:** distribuendo i dati in modo ridondante su più dischi fisici

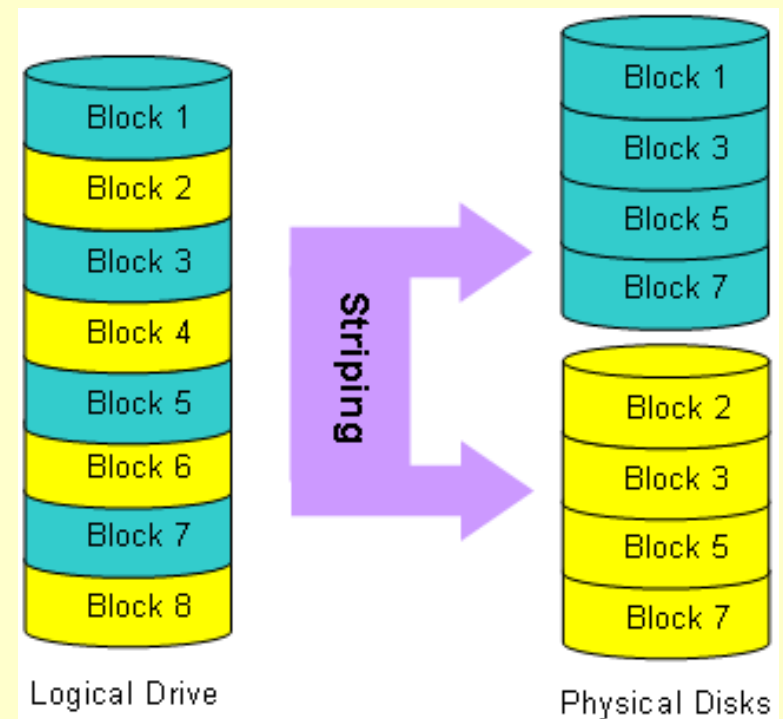
Il RAID appare al sistema operativo come un unico disco rigido

5. RAID: Striping (RAID 0)

In questa configurazione i dati vengono scritti a **strisce** su più dischi

- velocità di trasferimento molto elevata quindi alte prestazioni
- non supporta il fault tolerance (se uno dei dischi si rompe, l'effetto si ripercuote su tutto l'array)

La capacità dell'array è data dal numero dei dischi moltiplicato per la capacità più bassa tra essi.

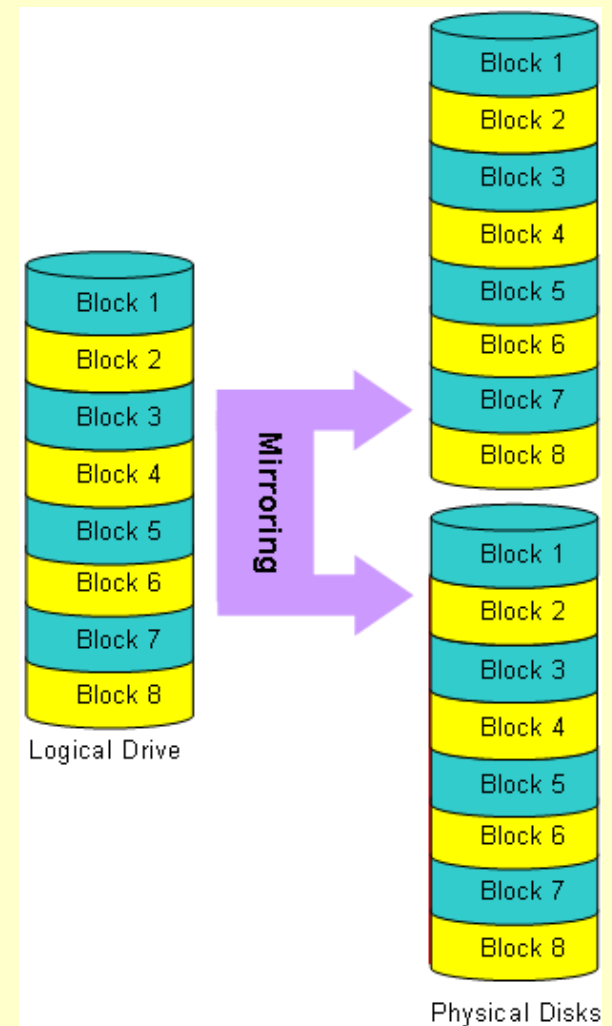


5. RAID: Mirroring (RAID 1)

RAID livello 1 usa almeno due dischi duplicati dove dispone esattamente gli stessi blocchi di dati

- velocità di trasferimento non molto elevata poiché i dati devono essere duplicati
- supporta il fault tolerance (se uno dei dischi si rompe, i dati sono recuperabili dall'altro disco)

La capacità dell'array è la metà della capacità totale dei dischi a causa della ridondanza

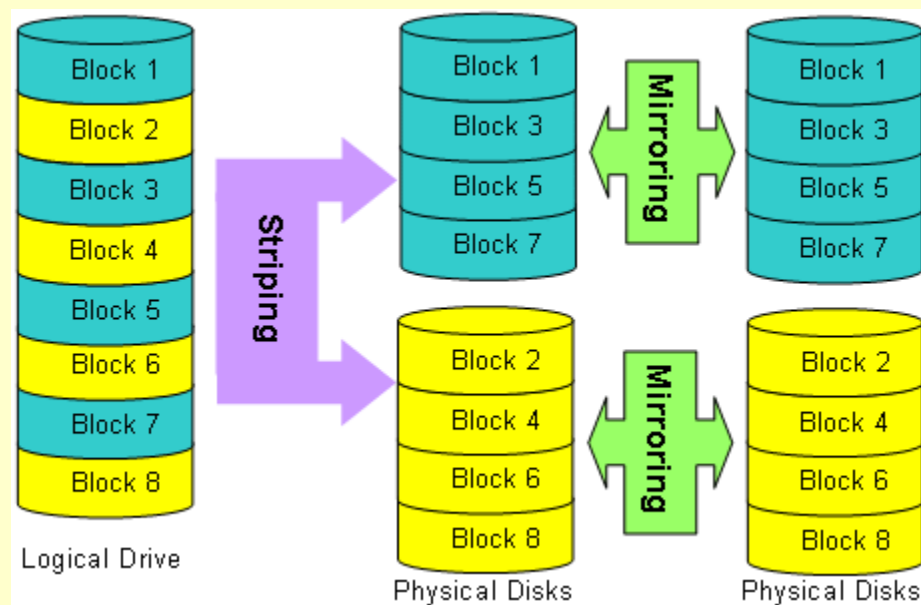


5. RAID: Striping con Mirroring (RAID 0 + 1)

Utilizza due dischi per memorizzare una striscia e li duplica su un altro per avere il fault-tolerance (sono necessari almeno 4 dischi)

- accesso veloce dei dati come RAID 0
- supporta il fault tolerance come RAID 1

La capacità dell'array è la metà della capacità totale dei dischi a causa della ridondanza



6. Schemi di rotazione dei supporti



Un aspetto fondamentale per una buona progettazione di un backup è prevedere uno ***schema di rotazione*** dei supporti per copiare i dati almeno una volta al giorno.

Un buon schema di rotazione deve garantire una *lunga, approfondita e svariata storiografia* di versioni dei file copiati per permetterne un buon ripristino

Schemi principali

- Grandfather - Father – Son (Best Practice)
- Tower of Hanoi (Ottimo)

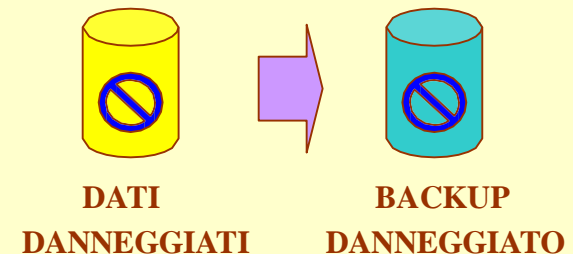
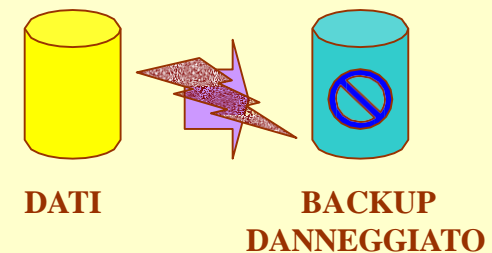
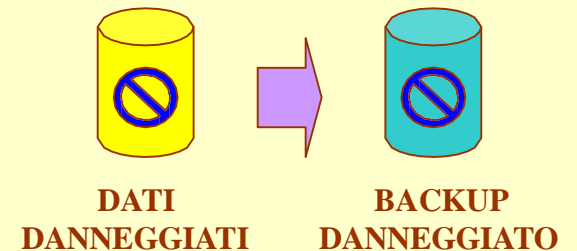
6. Motivazioni per schema di rotazione

- l'esecuzione del backup su un unico supporto

di un dato danneggiato **propaga** il danno al backup

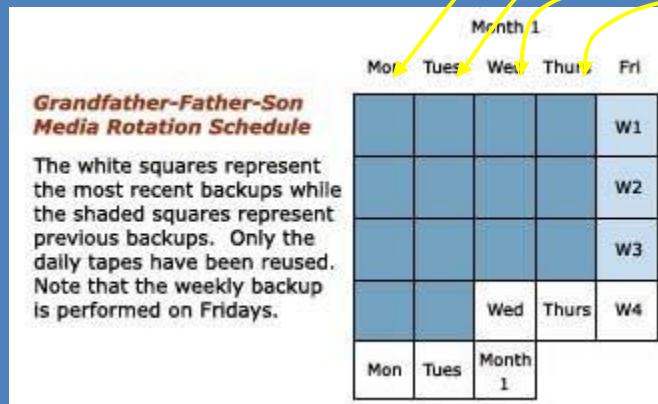
- se **durante la fase di backup** su un unico supporto si ha un evento dannoso, si danneggia l'unica copia di backup

- se dopo avere effettuato correttamente la fase di backup su un unico supporto, si ha un evento dannoso **durante il ripristino** si ha la completa perdita dei dati stessi



6. Schemi di rotazione: Grandfather-Father-Son

Questo schema utilizza insiemi di backup giornalieri (Son), settimanali (Father) e mensili (Grandfather).



Son 1

Son 2

Son 3

Son 4

Father 1

Father 2

Father 3

Father 4

Father 5

GrandFather 1

GrandFather 2

GrandFather 3

Per questo schema di rotazione sono necessari 12 supporti e permette una copertura di 3 mesi.

6.Schemi di rotazione: Tower of Hanoi

In questo schema ogni supporto di tipo A è usato alternandolo giornalmente.

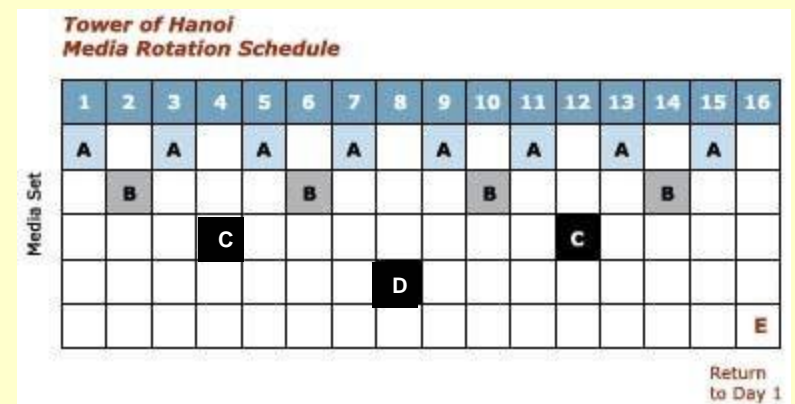
Il secondo tipo di supporto B è utilizzato il primo giorno libero dopo A e ripetuto ogni 4 sessioni di backup.

Il terzo tipo di supporto C è utilizzato il primo giorno libero dopo A e B e ripetuto ogni 8 sessioni di backup.

Il quarto tipo di supporto D è utilizzato il primo giorno libero dopo A , B e C e ripetuto ogni 16 sessioni di backup.

I supporti D ed E si alternano.

Con ogni supporto aggiuntivo
aggiunto allo schema di rotazione il
backup raddoppia le sue
informazioni



7. Best Practice sul backup



- I backup devono essere effettuati con cadenza almeno giornaliera.
- Il requisito minimo per la maggior parte delle organizzazioni è quello di effettuare un backup completo ogni settimana e backup incrementali ogni giorno.
- Almeno una volta al mese il supporto di backup deve essere verificato effettuando un ripristino su un server di prova per controllare che la procedura utilizzata sia corretta.
- Le soluzioni di backup di massimo livello sono costituite da reti totalmente ridondanti con capacità di recupero degli errori (per sistemi finanziari e di e-commerce che operano in tempo reale, per sistemi che controllano infrastrutture critiche e per alcuni sistemi militari).

7. Trattamento dei supporti di backup




- Conservare le copie di backup in luogo sicuro e al riparo da rischi.
- Conservare le copie di backup lontano dal luogo dove si trovano le macchine che contengono i dati copiati (se il computer viene rubato o si verifica un incendio, o un disastro anche la copia di backup viene persa).
- In generale utilizzare ogni giorno un disco o un supporto fisico diverso
- se si usa un cloud, aree geografiche o geopolitiche diverse

8. Recovery dei dati



- Forward error recovery
- Backward error recovery
- Checkpoint e recovery in sistemi networked
 - Checkpoint e recovery sincrono
 - Checkpoint e recovery asincrono

9. Operazione sui file di backup: Compressione



La dimensione dei file costituisce un problema sia per l'archiviazione sia per il trasferimento dei file stessi.

Comprimere un file significa ridurre le dimensioni applicando particolari ***algoritmi matematici***.

Esempi di compressioni sono:

- zip - rar
- gif -jpeg - tif
- mp3 - wav

9. L'efficienza del coefficiente

Coefficiente o fattore di compressione :

L'efficienza della compressione viene calcolata dividendo la grandezza originale del file per la sua grandezza una volta compresso (compression rate).

Lossless : formato che è in grado di restituire, al termine della decompressione, un'immagine esattamente uguale -pixel per pixel -all'originale com'era prima che venisse compresso.

Lossy : formato di compressione che non può assicurare una reversibilità assoluta (con perdita o distruttivo).

9. L'algoritmo Huffman

" o Tinto tu che hai ritinto il tetto non te ne intendi tanto" (di tetti ritinti)

numero di ricorrenze:

o numero ricorrenze 6

Il numero ricorrenze 1

d numero ricorrenze 1

t numero ricorrenze 12

i numero ricorrenze 7

n numero ricorrenze 8

u numero ricorrenze 1

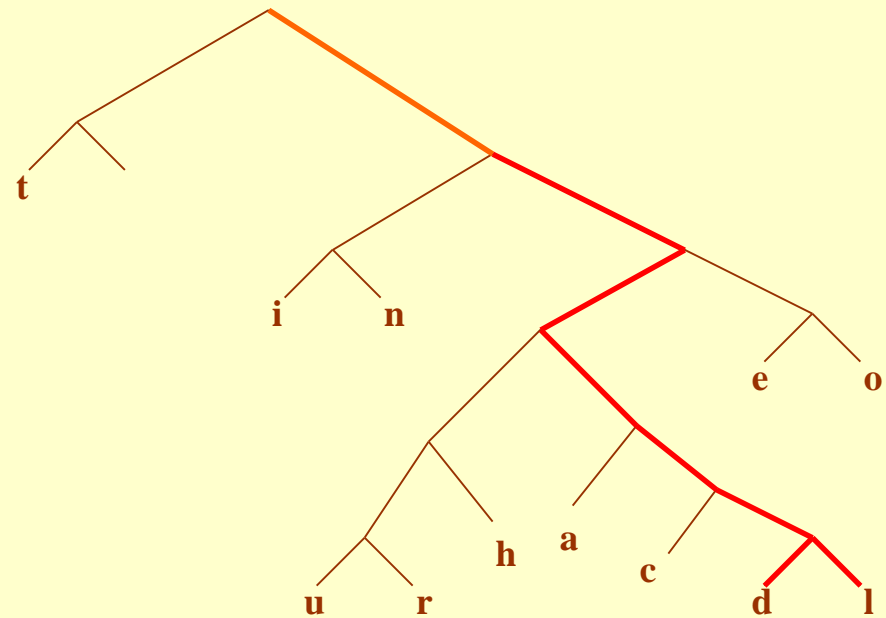
c numero ricorrenze 1

h numero ricorrenze 2

a numero ricorrenze 2

r numero ricorrenze 1

e numero ricorrenze 5



9. L'algoritmo Huffman



Viene generato il file compresso, sostituendo a ciascun elemento del file originale il relativo codice prodotto al termine della catena di associazioni basata sulla frequenza di quell'elemento nel documento di partenza.

Il guadagno di spazio al termine della compressione è dovuto al fatto che gli elementi che si ripetono frequentemente sono identificati da un codice breve, che occupa meno spazio di quanto ne occuperebbe la loro codifica normale.

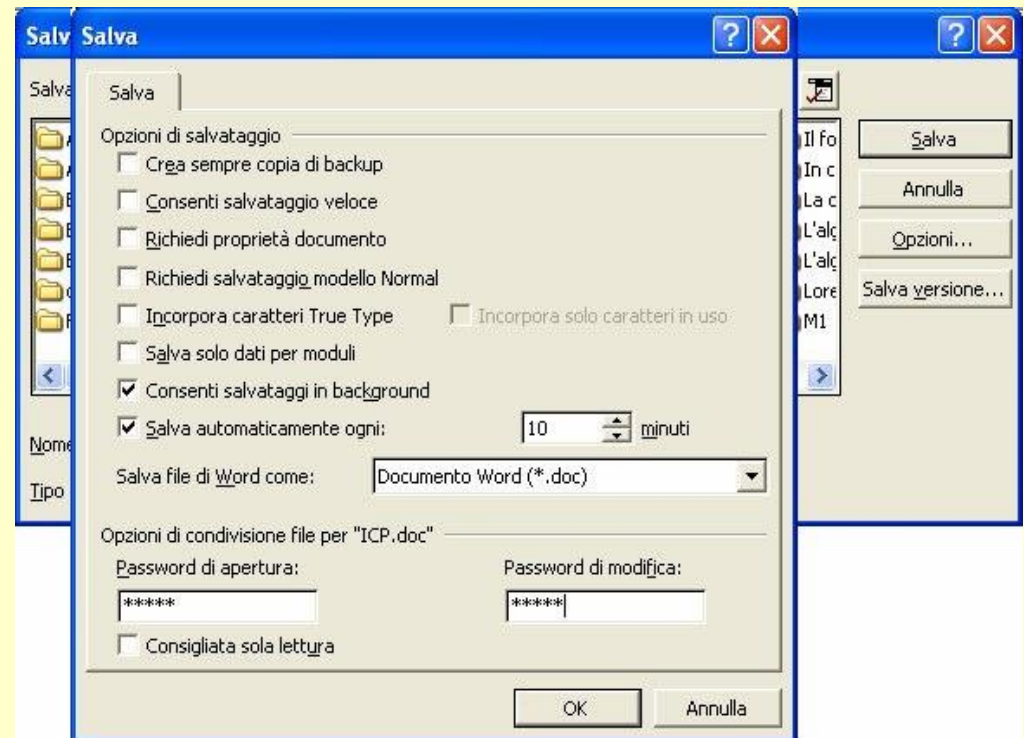
Viceversa gli elementi rari nel file originale ricevono nel file compresso una codifica lunga, che può richiedere, per ciascuno di essi, uno spazio anche notevolmente maggiore di quello occupato nel file non compresso.

Quindi questo tipo di compressione è tanto più efficace quanto più ampie sono le **differenze di frequenza** degli elementi che costituiscono il file originale, mentre scarsi sono i risultati che si ottengono quando la distribuzione degli elementi è uniforme.

9. Operazione sui file di backup: Password

I file protetti tramite password possono essere aperti solo dagli utenti autorizzati.

In genere i sistemi operativi includono un sistema di protezione tramite password e la maggior parte delle applicazioni software, fra cui Microsoft Office, consente di utilizzare le password per proteggere file e cartelle.



9. Operazione sui file di backup: Algoritmi hash

Funzione Hash: funzione che mappa un messaggio di lunghezza arbitraria in un valore hash di lunghezza fissa (128 o 160 bit), chiamata digest message o digital fingerprint ossia impronta digitale. Queste funzioni devono avere due proprietà:

-unidirezionalità, ossia dato x è facile calcolare $f(x)$, ma data $f(x)$ è computazionalmente difficile risalire a x .

-prive di collisioni (collision-free), ossia a due testi diversi deve essere computazionalmente impossibile che corrisponda la medesima impronta ($x \neq y$ allora $f(x) \neq f(y)$).

Principali algoritmi:

- MD4 (128 bit)
- MD5 (128 bit)
- SHA1 (160 bit)

9. Minacce: Password Cracking

Nei moderni sistemi le password non vengono MAI salvate mediante un algoritmo di cifratura ma mediante un algoritmo di **hashing** (un algoritmo di cifratura e' reversibile uno di hashing no)

Per **password cracking** si intende il tentativo di risalire ad una password a partire dal suo hash andando a calcolare l'hash di tutte le possibili combinazioni di caratteri alfanumerici e di punteggiatura per poi confrontarlo con l'hash stesso per vedere se c'è corrispondenza.

tecnica con vocabolario utilizza dizionari di parole comuni come password e tentativi basati su varianti di queste generate con grammatiche ad hoc

brute forcing, ossia provare ad accedere al sistema mediante ripetuti tentativi di login.

Quest'ultima tecnica è chiaramente da evitare perché sul sistema obiettivo vengono loggati tutti i tentativi.

Conclusioni



Backup

- pianificazione:
 - performance
 - schemi di rotazione e ridondanza
- trattamento fisico dei supporti
- backup remoto
- conservazione dati a lungo termine