

Backup e Restore di Classe Enterprise

Stefano Noferi

stefano.noferi@noze.it

Elisa Balatresi

elisa.balatresi@noze.it

Pisa, 19 Marzo 2009

Introduzione

Le procedure di Backup e Restore costituiscono un aspetto fondamentale nella gestione dei dati.

L'importanza del salvataggio dei dati non è mai sufficientemente sottolineata ed è l'unica garanzia contro guasti hardware irreparabili e contro virus che possono alterare il contenuto di archivi.

Le statistiche dimostrano che il 66% degli utenti di internet ha subito perdite dati.

L'organizzazione dello spazio di backup e la gestione dei sistemi di backup è un'operazione tanto importante quanto critica.

Terminologia

BACKUP: in letteratura viene fatta confusione tra il concetto di Archivio, Backup e Sistema di Fault Tolerance.

Il BACKUP è un'operazione è volta a DUPLICARE su differenti supporti di memoria delle informazioni digitali secondo una periodicità stabilita e conforme alle politiche aziendali.

Il processo di archiviazione è volto a creare una copia primaria dei dati, tipicamente da destinarsi ad un uso futuro, mentre il processo di backup crea due copie degli stessi dati. In tal senso, un sistema RAID non può essere assimilato ad un sistema di Backup (ho un'unica copia in un'unica locazione)

I Sistemi di Fault Tolerance partono dall'assunzione che un fault non debba provocare perdite di dati, mentre il Processo di Backup interviene per contrastare la perdita di dati successiva ad un fault.

Terminologia

RESTORE: operazione che permette di ripristinare uno stato precedente a partire dal Backup disponibile.

L'operazione di Restore consente:

- il ripristino di uno stato precedente in caso di perdita di dati: **Disaster Recovery**.
- il ripristino di specifici file che sono andati corrotti o sono stati cancellati accidentalmente.

Backup Method – Full Backup

Nel seguito sono indicati alcuni tra i principali metodi di Backup.

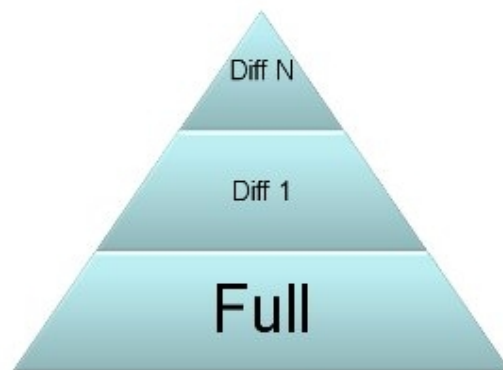
- **Full Backup**: backup dell'intero sistema (si differenzia dalla disk image perché non include: tabelle di allocazione, partizioni, settori di boot).
- **Disk Image**: metodo di backup di un intero disco o filesystem. Utile per ricreare copie esatte di un disco (può trovare applicazione nell'installazione aziendale su tanti computer di un medesimo ambiente).

Backup Method – Incremental Backup

Incremental Backup: consiste nella copia di tutti i file che hanno subito modifiche a partire dall'ultimo Full, o Incremental Backup. Esistono molteplici varianti di questo metodo

- vantaggi: maggior velocità rispetto al metodo del Differential Backup.
- svantaggi: tempi di restore (**RESTORE TIME**) più lunghi: per completare la procedura di restore è necessario partire dall'ultimo full backup ed aggiungere in sequenza tutti i backup incrementali.

Incremental Backup



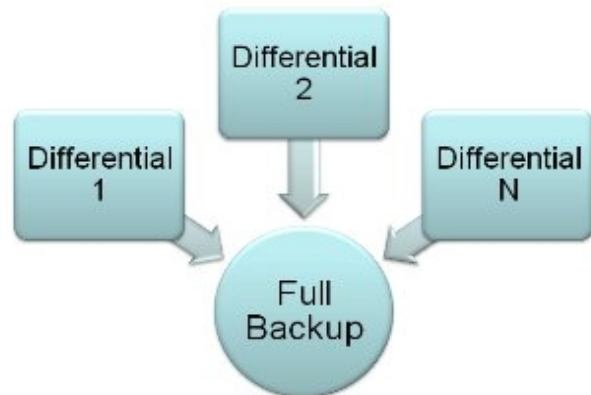
Restore with Incremental Method



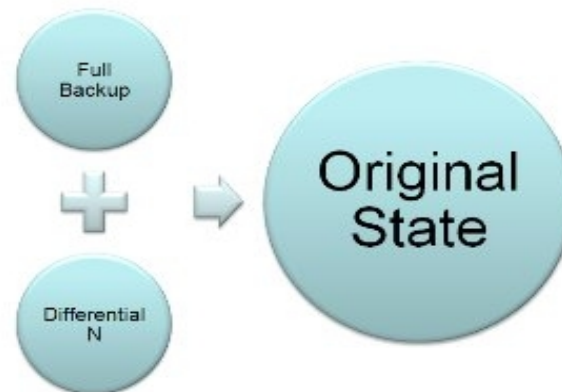
Backup Method – Differential Backup

- **Differential Backup**: è il backup cumulativo di tutti i cambiamenti effettuati a partire dall'ultimo Full Backup completo.
 - vantaggi: minor tempo nell'effettuare il backup rispetto al Full Backup.
 - vantaggi rispetto all'Incremental Backup: Il RESTORE TIME è data dalla somma del tempo di restore dell'ultimo full backup e di un singolo Differential Backup.
 - svantaggi: i dati aumentano ogni giorno che passa dall'ultimo Full Backup.

Differential Backup Method



Restore with Differential Method



Storage Media

La scelta e la gestione del media utilizzato per il backup è un'operazione tanto **importante** quanto **critica**.



Storage Media - Criticità

- La scelta dei data storage devices deve essere effettuata tenendo presenti aspetti di sicurezza, ridondanza, capacità di memorizzazione, velocità di trasferimento, costi ed affidabilità.
- I supporti su cui viene effettuato il backup devono essere periodicamente verificati o sostituiti. (alcuni parametri da considerare per hard disk: MTTF, temperatura interna, presenza sistemi RAID che modificano il valore MTTF di un singolo disco).
- Deve essere prevista la dislocazione dei supporti di backup in posizioni distinte e separate dai sistemi in uso al fine di prevenire furti, incendi, alluvioni e/o altri eventi catastrofici.

Software di Backup – Criteri di valutazione

La scelta di un software di backup piuttosto che un altro è legata alle funzionalità che offre rapportate alle specifiche esigenze. Di seguito sono elencate una serie di caratteristiche peculiari di un software di Backup :

- **Copia immagine di un disco rigido.**
- **Copia selettiva di directory e singoli file.**
- **Criteri di selezione** per la ricerca dei contenuti backup, e per la scelta di quelli che devono essere oggetto di backup.
- **Portabilità**: ad esempio la possibilità di effettuare il Backup su una distribuzione Linux e poi effettuare il Restore su un'altra ('dd', 'cpio' 'tar', 'dump').
- **Multiplexing**: offre la possibilità di effettuare backup simultanei su un singolo dispositivo di archiviazione.
- **Compressione**: permette di ridurre lo spazio di backup.

Software di Backup – Criteri di valutazione

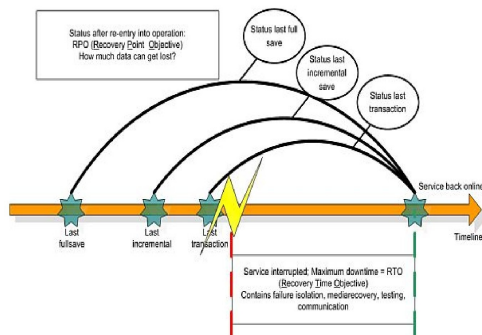
- **Algoritmi di De-duplicazione**: questa funzionalità consente di eliminare i duplicati all'interno degli Storage Media utilizzati per il Backup.
 - L'operazione di De-duplicazione Può essere eseguita prima, durante, dopo il backup, in contemporanea o in differita rispetto all'operatività dei sistemi informativi.
 - Particolarmente importante per file di grandi dimensioni (es database) o in ambienti in cui si utilizzano file condivisi da molti utenti.
 - Riduzione dello spazio di backup richiesto (dato meno critico in quanto negli ultimi anni il costo delle memorie informatiche è diminuito secondo la Legge di Moore)
 - **Incremento della velocità di backup e di restore**. Dato critico per le aziende, che devono consentire il completamento delle attività di backup durante la notte (FINESTRA DI BACKUP), in modo che queste non vadano ad interferire con l'operatività quotidiana introducendo un sovraccarico dei sistemi informatici ed il rallentamento delle attività degli utenti).

Software di Backup – Criteri di valutazione

- **Sicurezza** in termini di protezione dei dati con password ed algoritmi di crittografia. Il processo di Encryption introduce alcune importanti criticità:
 - costituisce un processo ad alto consumo di CPU che può ridurre pesantemente la velocità di Backup.
 - i dati cifrati non possono essere compressi. Per tale motivo molti algoritmi di encryption provvedono a comprimere i dati in prima di cifrarli.
 - le politiche di gestione delle chiavi devono essere accurate per non inibire la sicurezza introdotta dagli algoritmi di encryption.
- **Accesso via rete.**
- **Automatismi.**
- **Interfaccia user-friendly.**

Processo di Backup e Restore – Management

La progettazione della strategia di Backup e Restore richiede la valutazione e delinearne, nel rispetto delle specifiche politiche aziendali di alcuni parametri peculiari.



- **Recovery Point Objective (RPO):** Rappresenta il massimo tempo tra la produzione di un dato e la sua messa in sicurezza (ad esempio attraverso backup) e, conseguentemente, fornisce la misura della massima quantità di dati che il sistema può perdere a causa di guasto improvviso.
- **Recovery Time Objective (RTO):** Rappresenta il tempo necessario per il completo recupero dell'operatività di un sistema a seguito di fault. (Massimo tempo di downtime).
- **Data Security:** è necessario utilizzare sistemi di Backup che garantiscano contro accessi non autorizzati.

Processo di Backup e Restore – Management

Al contempo la progettazione del sistema di Backup e Restore deve tenere conto di alcuni limiti:

- **Backup window**: periodo di tempo in cui un sistema è disponibile per il backup. La Backup window deve essere progettata in modo da avere la minima interferenza possibile con l'operatività quotidiana dei sistemi informativi.
- **Performance impact**: Il processo di Backup introduce un impatto sulle performance che deve essere tenuto presente durante la sua progettazione.
- **Costs of hardware, software, labor**: La scelta oculata del media di backup in relazione allo spazio di backup di cui si necessita rappresenta un punto importante nella progettazione del processo di Backup.

Network Bandwidth: Questo parametro ha un impatto importante sulle performance del processo di Backup e Restore.

I principali software di Back-up Open Source

- BackupPC: <http://backuppc.sourceforge.net/>
- Bacula: <http://www.bacula.org/en/>
- Amanda/zamanda: <http://www.amanda.org/>

Questo Software ha tra le sue peculiarità quella di consentire un backup in rete dei dati presenti su workstation e server, indipendentemente dal sistema operativo utilizzato, sia esso Linux, Unix o Microsoft Windows. Offre la possibilità di creare dei backup incrementali, oppure effettuarli su più volumi (in modo da non dover cambiare tecnologia qualora la dimensione dei nostri dati superi quella del supporto di backup) e degli ottimi strumenti per la compressione.

- dump: <http://sourceforge.net/projects/dump/>
- rsync: <http://rsync.samba.org/>
- tar: <http://www.gnu.org/software/tar/>

Grazie

Stefano Noferi

stefano.noferi@noze.it

Elisa Balatresi

elisa.balatresi@noze.it

Pisa, 19 Marzo 2009