

Copy On Write Filesystem come strumenti di backup incrementale

Mattia Valzelli

Università degli studi di Verona

11 Luglio 2013

Perché studiare i filesystem moderni

Le misure per evitare la perdita dei dati anche dopo un evento distruttivo si sono evolute nel tempo e affinate. Queste tecniche però si concentrano su problematiche hardware nella perdita dei dati, trascurando la parte relativa ad errori del software o della persona fisica (cancellazione dati per errore, corruzione dati etc..).

La maggior parte dei filesystem moderni non offre metodi efficaci e completi per evitare questi danni all'utente comune.

Cosa è un filesystem

Un filesystem è la parte di un sistema operativo dedicata alla gestione delle informazioni su disco (i file).

Dal punto di vista dell'utente viene astratto:

- come le informazioni vengono esposte (gerarchia, directory);
- regole di visualizzazione, modifica (permessi);

Un filesystem ha il compito di nascondere tutte queste problematiche.

Questi software sono stati creati quando le prestazioni e le dimensioni dei dischi erano totalmente differenti da quelli del giorno d'oggi. Inoltre, negli ultimi anni stanno emergendo tipologie di dischi che funzionano meccanicamente in modo totalmente diverso (SSD).
I filesystem moderni inoltre non offrono nessuna funzionalità "nativa" per creare dei backup.

Microsoft

Microsoft propone una utility chiamata "Ripristino configurazione di sistema" che cerca di riportare il sistema operativo, file di sistema, chiavi di registro e/o programmi installati ad un punto preciso nel tempo, nel quale l'utente ha la sicurezza che il sistema funzioni correttamente o comunque con un comportamento coerente alle aspettative.

Da Windows Vista e successivi Microsoft ha invece sviluppato un software separato per la gestione di questa funzionalità chiamato "Shadow Copy". Questa nuova modalità per il ripristino ha il vantaggio di poter monitorare e copiare per backup i file a livello di filesystem block di un qualsiasi file su disco.

Ci sono svantaggi:

- backup viene configurato per tutto il volume;
- mancanza di possibilità di selezione di quali file includere;
- backup memorizzato sul volume stesso.

Linux

Non è mai emerso uno strumento efficace e comunemente usato come standard se non per la utility di sistema `dd(1)`. Essa è un programma disponibile da Unix Version 7, che copia per blocchi di byte da un input ad un output. Diventa quindi possibile creare script per effettuare backup manuali e periodici.

Ubuntu, la distribuzione più diffusa e user friendly, dalla versione 11.04 (Aprile 2011) ha incluso uno strumento di backup chiamato Dèjà Dup poi successivamente rinominato Backup.

Backup crea un dump completo, mentre le successive versioni sono incrementali (per queste versioni incrementali viene usato il programma rdiff).

Purtroppo, come accennato precedentemente, non è disponibile nessun software standard per il backup completo del sistema; ma soprattutto non esiste niente di accessibile che possa aiutare l'utente base ad interagire in modo semplice ed intuitivo.

Apple

Time Machine crea backup di tutto il sistema operativo, compresi i file personali dell'utente, in maniera incrementale.

Questo backup deve essere creato su un supporto diverso da quello di partenza, sia che esso sia un disco esterno o un'altra partizione dello stesso.

Time Machine utilizza dei link simbolici per effettuare backup incrementali. Questo è positivo da un lato perché permette di risparmiare spazio disco, ma dall'altro riduce il punto di fallimento a un solo file. Time Machine effettua un backup a livello di file, ossia ogni volta che parte un servizio per il backup ottiene la lista dei file modificati e ne crea una nuova versione sul supporto di destinazione.

È ciò che più si avvicina a quello che cerchiamo.
Facile da usare per l'utente, grafica accattivante, possibilità di effettuare dei backup su un disco di rete.

Problema comune a tutte le proposte analizzate

Le tecniche descritte hanno tutte dei problemi comuni:

- scarsa possibilità di configurazione;
- il device su cui effettuare il backup è interno o esterno;
- nessuna tecnica permette di verificare che i dati salvati non siano danneggiati.

ZFS

ZFS è un filesystem di nuova generazione sviluppato da Sun Microsystems (ora Oracle) ed è considerato un innovativo, completo ridisegno di un filesystem UNIX tradizionale.

Tra le features più rilevanti troviamo:

- Protezione contro la corruzione dei dati;
- Supporto per la memorizzazione di una grande quantità di dati;
- Integrazione del concetto di filesystem e di gestione volumi;
- snapshot;
- Copy-On-Write cloni;
- Controllo continuo della validità dei dati e riparazione automatica;
- RAID-Z;
- NFSv4 ACL native.

A differenza dei filesystem tradizionali, che risiedono su un device (volume) fisico singolo e quindi richiedono un volume manager per usare più di un device, i filesystem ZFS sono costruiti sopra un layer di storage pools virtuali chiamati zpool.

Uno zpool è composto da virtual device vdevs, che sono loro stessi costruiti su block device: file, partizioni di hard drive, interi drive, con quest'ultima modalità indicata come preferita.

ZFS usa un modello transazionale Copy-On-Write. Tutti i blocchi puntatori nel filesystem contengono un checksum a 256 bit oppure un hash a 256 bit (a scelta fra Fletcher-2, Fletcher-4 o SHA-256) del blocco a cui punta che viene verificato quando viene letto. Blocchi che contengono dati attivi non sono mai sovrascritti sullo stesso; invece ne viene allocato uno nuovo, le informazioni modificate vengono scritte in questo nuovo spazio, poi tutti i metadati che fanno riferimento ai dati vecchi vengono letti, riallocati e riscritti nello stesso modo.

Snapshot e Clone

Il modello Copy-On-Write utilizzato da ZFS ha un altro grande vantaggio: quando ZFS scrive nuovi dati preserva i blocchi contenenti i vecchi invece di liberarli o sovrascriverli, creando una versione "snapshot" del filesystem.

Cloni sono un tipo particolare di snapshot, in quanto sono modificabili. Infatti alla creazione di un clone viene istanziato un filesystem indipendente che comunque condivide i blocchi con il filesystem "padre". Modificando il filesystem clone nuovi blocchi dati sono creati per riflettere questi cambiamenti, ma blocchi dati non modificati continuano ad essere condivisi (a prescindere da quanti cloni esistano).

ZFS send

Il comando `zfs send` crea uno stream di byte rappresentante uno snapshot che viene scritto su standard output. Di default uno stream completo viene generato. Come un qualsiasi stream esso può essere reindirizzato su un file o addirittura su un altro sistema.

Per creare uno stream di byte non completo ma incrementale basta specificare l'opzione `-i`.

ZFS receive

Il comando ZFS receive è la controparte di ZFS send. Esso permette di creare uno snapshot i cui contenuti sono specificati dallo stream che è fornito nello standard input. Se in ingresso è fornito un filesystem completo ne viene creato uno nuovo che lo rappresenta.

Combinare send e receive

Questi due comandi sono il cuore per poter automatizzare una strategia di backup. Infatti grazie a ZFS send possiamo trasmettere lo stato corrente del nostro filesystem, con ZFS receive possiamo salvarlo e riutilizzarlo successivamente.

Questo traffico tra send e receive può essere effettuato tramite un qualsiasi protocollo per lo scambio di dati: infatti dobbiamo solo avere un protocollo che permetta uno stream di dati.

I punti deboli nei sistemi citati

Nei sistemi analizzati abbiamo citato i seguenti problemi:

- backup viene configurato per tutto il volume;
- scarsa possibilità di selezione riguardo i file da includere;
- backup memorizzato sul volume stesso.

ZFS ovvia tutti questi problemi:

- backup può essere configurato per filesystem o dataset (creabili e definibili dall'utente);
- creando una struttura di dataset in modo opportuno si configura che file includere;
- backup memorizzato dove desidera l'utente (snapshot possono essere tenuti sullo stesso disco oppure mandati ad un altro su cui è stata salvata una versione precedente del filesystem).