Backup Handbook

Andrea Grazioso
<grazioandre at gmail dot com>

2014

Contents

Premessa

Questo Handbook è stato fatto mentre preparavo il talk riguardante i backup per i Corsi Linux Avanzati (edizione 2013/2014) tenuti dal POuL al Politecnico di Milano. Realizzati di pari passo alle slides utilizzate tra notti insonni e feroci ricerche su libri e pagine web sono diventati un bel malloppo di cose più o meno utili che potrebbero servire all'occorrenza per rinfrescare qualche concetto.

Potrebbero essere presenti errori come imprecisioni, quindi non do nessuna garanzia sulla correttezza dei contenuti e non mi assumo nessuna responsabilità in termini di danni arrecati a cose/persone/computer nel caso di uso improprio dei comandi riportati nella sezione apposita. Sono gradite le segnalazioni di errori.

Cos'è il backup? A cosa serve?

`There are two kinds of people: those who do regular backups and those who never had a hard drive failure - Unknown.`

Il motivo per cui esiste il backup è semplicemente dovuto al fatto che il contenuto dei nostri pc dopo qualche tempo di utilizzo arriva a valere più del PC stesso. Con il termine backup ci si riferisce all'atto di copiare e archiviare su un device diverso da quello su cui si opera normalmente cosicché sia possibile un eventuale

ripristino dell'intero sistema o di singoli file in caso di incidenti, erronee cancellazioni, malfunzionamento di un hard disk, furto, corruzione dei dati e altre sventure più o meno creative.

Tutto il sistema su cui si lavora può essere incluso nel backup, vedremo tecniche che includeranno informazioni più o meno dettagliate sullo stato del sistema o che opereranno sui singoli file. Per i sistemisti VPS/server è fondamentale conservare i backup del sistema per prevenire qualsiasi evenienza ed eventualmente proteggere gli utilizzatori dei propri servizi da perdite di dati.

A questo proposito è importante ricordare che qualsiasi drive verrà scelto come destinazione di backup è caldamente consigliato che sia "relativamente" nuovo per assicurarsi che sia allo stesso tempo esente da difetti di fabbrica e da difetti legati all'usura, ad esempio le chiavette usb esauriscono la loro vita dopo approssimativamente 100'000 cicli di scrittura in modo imprevedibile e improvviso, e di sicuro a nessuno farebbe piacere dopo la morte di un hard disk constatare che anche l'unità su cui sono i backup è illegibile. Stesso discorso vale per gli hard disk, che sono più prevedibili in quel senso, ma vanno comunque incontro a problemi di vecchiaia come il fallimento di settori, o, molto più grave, guasto della testina di lettura. Inoltre è molto importante scegliere un posto fisico in cui locare il drive o i drive contenti le copia di backup, che è bene rimangano confinati ad un unica funzione, cioè quella di backup; dopo aver assolto questa funzione il drive in questione dovrebbe essere rimosso e collocato in un posto sicuro, lontano dal PC e da mani inaffidabili.

In caso si potrebbe anche optare per il backup via rete, possedendo un NAS (Network Attached Storage), un server abbastanza capiente o un account cloud supportato dalle utility di backup.

L'unica accortezza è che in caso di crash del sistema il backup dovrebbe essere facilmente accessibile per il ripristino (quindi niente password complicate memorizzate solo nel dispositivo fallito), ma in quanto copia del sistema rappresenta un dato sensibile quindi attenzione a metterlo online sprovvisto di qualsivoglia forma di crittografia.

Supporti di backup

A partire dai floppy disk ai più moderni SSD qualsiasi supporto in grado di memorizzare informazioni si presta per effettuare copie di backup dei propri file, ovviamente alcuni supporti saranno più indicati di altri per durata nel tempo, capacità e velocità di trasferimento.

Nastri Magnetici

Utilizzati principalmente in ambienti dove l'affidabilità dei supporti è un requisito fondamentale, come nelle banche, nelle grosse corporazioni che trattano giornalmente dati di decine di migliaia di fornitori/dipendenti/clienti e una perdita di dati significherebbe la perdita di grosse quantità di denaro. Ovviamente per scrivere su di un nastro magnetico vi è la necessità di possedere appositi macchinari costosi ed ingrombanti, accessibili eclusivamente alle sopracitate categorie. Si stima che la durata massima dei nastri magnetici migliori tenuti in perfette condizioni di umidità/temperatura sia di circa 50 anni e ad oggi i nastri magnetici più performanti sul mercato hanno una capacità di archiviazione di circa una decina di TB e un rate di trasferimento di 250 MB/s.

Pro

- Ottima longevità
- Grande capacità di archiviazione
- Le unità a nastro più recenti permettono di effettuare la compressione dei file "on the fly"
- Ottimo rate di trasferimento

Contro

- Necessitano di unità a nastro...
- ...molto ingombranti...
- ...e costose

Floppy Disk & Zip

I Floppy Disk sono supporti magnetici in cui l'informazione viene salvata tramite magnetizzazione. Vengono utilizzati tramite appositi drive. Utilizzati principalmente negli anni '80 e '90 sono oramai deprecati.

Pro

- Poco Costosi
- Molto Portatili
- Durata media in condizioni ottimali(10 anni)

Contro

- Dimensione oggi insignificante 1.44 megabytes (sufficiente a contenere "solo" 10 9 caratteri di testo standard)
- Richiedono un drive esterno per essere utilizzati
- Bassa velocità di trasferimento
- Eccessiva delicatezza

I "Zip" Poco noti per via dell'avvento del CD gli zip rappresentano la "seconda generazione" di floppy disk, hanno dimensioni fisiche simili ai floppy drive classici ma capacità maggiori (100 MB, 250 MB e 750 MB) e durano circa una decade. Oramai anche questo tipo di supporti sono deprecati e fuori commercio quindi passeremo oltre.

Dispositivi Ottici

CD, DVD e Blu-ray rappresentano un buon supporto per il backup. Sono formati da strati sovrapposti di alluminio in cui i dati sono incisi tramite laser formando una trama di microfori che viene successivamente letta in modo analogo. Invece le controparti riscrivibili di questi supporti sono formati da un composto di metalli pesanti sensibili al calore che permette di essere forato e riamalgamato a seconda delle esigenze. Questi drive sono in circolazione da molto tempo in quanto hanno spopolato nel mercato diventando il principale mezzo di diffusione di materiale audio, video, software e videoludico, possiedono una moderata longevità: 10 anni per i CD, 30 per i DVD, 50 per i Blu-ray; ma spesso questi dati sono imprecisi, infatti questi supporti sono notevolmente influenzati dal modo in cui sono stati prodotti (attenzione al retro della confezione), con quali materiali e addirittura uno stesso modello può avere caratteristiche diverse a seconda della fabbrica in cui è stato prodotto. Inoltre vi sono altri fattori che caratterizzano questo supporto, quali la necessità di possedere un drive apposito per poterli utilizzare. Degna di nota la loro impermeabilità e la resistenza al manetismo, e ai lievi danni superficiali grazie ai moduli di controllo degli errori. Una delle principali peculiarità di questi supporti è la possibilità di effettuare una singola scrittura (CD-R, DVD-R, DVD+R per citarne alcuni). Nonostante possa sembrare un grande svantaggio invece si può rivelare un ottima soluzione in caso si voglia essere certi che i dati scritti non siano più manipolabili o eliminabili accidentalmente. Altri formati di CD, DVD e Blu-ray sono riscrivibili, caratteristica limitata a poche centinaia di riscritture.

Pro

- Dimensioni variabili (700MB per i CD, 4.7GB per i DVD, 200GB per i Blu-ray)
- Longevità medio-alta (10 anni per i CD, 30 per i DVD, 50 per i Blu-ray)
- Facilità di consevazione
- Resistenza ad alcuni tipi di danni
- Si può optare per vari formati
- Per applicazioni specifiche è conveniente la non riscrivibilità di una parte di questi supporti

Contro

- Necessitano di un drive apposito per la scrittura
- Sensibili ai graffi
- La tecnologia con cui si legge/scrive è negativamente influenzata da vi- brazioni e movimenti

Riassumento

Device	Max Dim.	Lifetime	Min Wrt Speed	Max Wrt Speed
CD	700MB	10	1x=150Kib/s	56x=8400KiB/s
DVD	4.7GB	30	1x=1.32Mib/s	24x=32.46MiB/s
Blu-ray	200GB	50	1x=4.29Mib/s	16x=68.66MiB/s

USB Flash Stick

Questi supporti negli ultimi anni sono diventate estremamente popolari per comodità, portabilità e riutilizzabilità (circa 100'000 cicli di scrittura). Buone per contenere quantità di dati variabili, ma principalmente per scambio di file di medie-piccole dimensioni, a fine 2013 sono arrivate a raggiungere capienza massima di 1 TB, possiedono una durata di vita di circa 10 anni, ma questo numero è fortemente influenzato dal tipo di controllore e chip di memoria installato, nonchè dall'utilizzo del drive stesso. Sono mediamente più veloci, piccole e capienti dei dispositivi ottici e queste memorie hanno il grande vantaggio di essere prive di parti in movimento, infatti questo tipo di memorie sono completamente elettroniche. Ma se ciò da un lato può sembrare un grande vantaggio non lo è in termini di affidabilità: le chiavette USB infatti sono note per essere soggette a corruzione dei dati se non correttamente maneggiate, inoltre è bene sapere che in caso di "morte" del device è impossibile recuperare i dati al suo interno. Le velocità dichiarate dagli standard USB sono per la versione 2.0 60MB/s e 3.0 625MB/s anche se il throughtput effettivo della trasmissione è rispettivamente di 35 MB/s e 200 MB/s al massimo per via dell'inadeguatezza del protocollo a supportare tali velocità, e dei vari overhead che ingombrano i trasferimenti. Infine un altro motivo per non consigliarle come device di backup è il costo elevato di questi dispositivi che incrementa notevolmente man mano che si richiede una maggiore capacità.

Pro

- Possono contenere una gran quantità di informazioni
- Compatibili con l'ormai diffusissimo standard USB
- Buona velocità di trasferimento (vedi tabella sotto)
- Estrema Versatilità e Portabilità
- Molto resistenti agli urti
- Buona longevità

Contro

- Il costo per GB è piùttosto elevato
- Facili da perdere
- Se non correttamente maneggiate si rischia di corrompere i dati al loro interno

Note: In realtà lo standard USB a partire dalla versione 1.0 è stato più volte ridefinito, infatti è stato inizialmente pensato per il collegamento di periferiche di input come mouse e tastiera che sono per definizione a bassissimo bitrate, per cui la velocità iniziale delle prime porte era molto bassa ed è andata aumentando via via nel tempo per adattarsi alle nuove tecnologie.

Hard Disk

Questi supporti sono formati al loro interno da dischi magnetici in rapida rotazione attorno al loro asse comune, letti da una testina magnetica mossa da un attuatore. Negli ultimi anni gli Hard Disk sono diventati il primario sistema di memoria di massa nei PC, per via della grande capacità, della buona velocità di trasferimento dati, e dal modico costo. Inoltre la loro capacità massima è il continua espansione, e nonostante l'introduzione dei ben più veloci SSD continueranno ad essere per molto tempo il principale sistema per archiviare grandi quantità di informazioni. Ad oggi (2014) il costo di un gigabyte negli hard disk è molto basso, si attesta circa a $0.05/GB, e la durata media è generalmente alta determinata quasi esclusivamente

dall'utilizzo, inoltre sono molto versatili, infatti possono essere connessi al PC tramite più interfaccie, PATA, SATA, Firewire, eSATA, USB, possono anche essere utilizzati da remoto tramite NAS o NFS. A loro sfavore ci sono la loro bassa resistenza a urti e lesioni in generale, infatti essendo costituiti da una tecnologia ibrida (meccanica e magnetica) sono molto sensibli agli urti e agli spostamenti bruschi come ai campi magnetici forti.

Pro

- Grande capacità
- Buona velocità di trasmissione
- Compatibili con qualsiasi interfaccia interna o esterna al pc
- Non necessitano di drivers per l'uso
- I materiali con cui sono costituiti hanno grande durabilità

Contro

- Necessitano stabilità durante l'uso
- Per essere utilizzati al meglio necessitano di un minimo di conoscenze sui filesystem

SSD

L' origine di queste memorie risale al 1950 ma la loro diffusione al mercato dei consumatori è avvenuta solo negli ultimi anni. Composti da circuiti integrati che fungono da memoria per l'archiviazione dei dati in questi dispositivi non c'è nessuna parte in movimento, il che li rende resistenti sotto ogni punto di vista, cioè permette di avere delle velocità irraggiungibili con qualsiasi altro dispositivo di memorizzazione di massa (i moderni SSD arrivano a superare i 500 MB/s in lettura e scrittura). La loro principale forza, cioè l'essere composti da celle di memoria di tipo elettronico è anche la loro debolezza, infatti sono soggetti a guasti irreparabili e improvvisi, perciò i tradizionali hdd sono preferiti nel caso si debba conservare una grande quantità di informazioni senza particolari esigenze di performance. Nel caso di VPS/server dedicati essi vengono utilizzati principalmente per assicurarsi le massime prestazioni dalla macchina ma più che altro per file di cache (in questo caso si parla di SSD cached), file temporanei ad accesso casuale o comunque file contenenti informazioni non critiche, altrimenti si ricorre sempre alla ridondanza dei dati tramite RAID/ZFS.

Pro

- Grande capienza
- Alta velocità di trasmissione dati
- Resistenti agli urti e ai movimenti bruschi
- Molto leggeri
- Molto silenziosi

Contro

- Costo elevato
- Bassa aspettativa di vita

Backup Online

Piùttosto che occuparsi personalmente del device su cui effettuare il backup si potrebbe optare per affidarsi a servizi di archiviazione cloud (Dropbox, Amazon Glacier, OVH Cloud Storage, Google Drive, etc). Tenere i backup in remoto aiuta a proteggerli da danni materiali quali furti, sequestri, incendi e in generale tutti i provider che offrono servizi di backup si occupano della ridondanza del dati e della loro protezione. In questo caso è bene assicurarsi che i dati messi in rete siano cifrati. Inoltre alcuni servizi di archiviazione offrono un certo grado di scalabilità, cioè la possibilità di acquistare spazio aggiuntivo che andrà quindi ad ampliare la quantità di spazio a nostra disposizione, e questo è senza dubbio più facile e veloce rispetto al dover acquistare un nuovo drive e occuparsi del passaggio dei dati da un drive all'altro.

Metodi di Backup

Quì entra in gioco una prima distinzione fondamentale, quella tra Snapshot e Backup, il primo più ingombrante e lungo da effettuare ma più facile da utilizzare e pronto all'uso in caso di crash del sistema, il secondo ottimizzato a seconda delle necessità, permette il versionamento dei file da salvare, il recupero selettivo, e i tempi di utilizzo delle risorse di sistema per effettuare i backup sono ridotti al minimo mantenendo alta la sicurezza.

Un immagine completa di sistema copia tutto bit per bit dal disco e quindi necessita di un dispositivo di destinazione più grande di quello da copiare. Viene copiato tutto, dalla tavola delle partizioni fino alla fine del disco, indipendentemente dal fatto che lo spazio sia effettivamente utilizzato o meno. Per fare ciò si usa dd

Invece con un backup non si ricrea la complessità del sistema operativo, ma si mira al salvataggio della struttura delle cartelle, dei file e tutti i loro dati e metadati. I tool usati più di frequente per lo scopo sono:

- rsync
- duplicity
- rdiff-backup

In generale possiamo distinguere tra due differenti approcci al backup:

- Incrementale, aggiunge i che hanno subito cambiamenti rispetto all'ultimo backup. Molto versatile ma necessità di un po di pratica per padroneggiarlo al meglio.
- Differenziale, modifica di volta in volta il backup precedente con i file che hanno subito dei cambiamenti. Più veloce dell'incrementale e meno ingombrante, ma mantiene solo l'ultima versione dei file.

Preparazione al backup

Arrivati a questo punto abbiamo un valido motivo per effettuare il backup di sistema, abbiamo il nostro device pronto per essere scritto, e abbiamo un idea di che tipo di backup vogliamo. Tipicamente un utente medio non dovrebbe preoccuparsi di altre directory se non la propria home, quindi potrebbe essere sufficiente effettuare un backup dei file nella home. In caso di backup completo al fine di ridurre lo spazio necessitato per la copia è utile smontare tutte le unità esterne (eccetto quella su cui andremo a scrivere la copia, ovviamente), svuotare il cestino, eliminare eventuali file temporanei lasciati in giro e chiudere quante più applicazioni possibili che possano scrivere su disco mentre il backup è in corso. Al fine di ridurre lo spazio necessario per il backup ci viene in aiuto un utile tool: ncdu. Questo software molto leggero utilizzabile comodamente da terminale permette di listare ricordivamente il contenuto di una directori (ad esempio la home) e riporta tramite un interfaccia grafica molto elementare come è ripartito il consumo della memoria per cartella così da poter individuare più facilmente eventuali sprechi di memoria.

Va tenuta in considerazione la struttura dei filesystem su Linux:

- /tmp è riservata ai file temporanei
- /media e /mnt contengono i percorsi di mount di periferiche esterne quindi non devono essere inclusi nel processo di backup
- /sys, /proc, /dev sono riferimenti virtuali alle periferiche e ai processi del sistema
- /var contiene i log, le cache
- /etc contiene principalmente file di configurazione predefiniti per le applicazioni. Utile per i backup del server, tipicamente no per quello dei desktop.

Programmi

Quì si apre veramente un mondo di scelte e possibilità, sostanzialmente dettate dalle proprie esigenze: sono disponibili sia tool da terminale, che di terze parti (solitamente i software di terze parti che fanno backup

fanno anche gestione di filesystem a livello più sofisticato ed hanno molte funzioni accessorie). E' bene fare attenzione al tool che si usa per trattare i propri dati, infatti non tutti i programmi sono in grado di ricreare la complessità di un sistema operativo e bisogna accertarsi personalmente che alcuni attributi di file e directory vengano rispettate, mi riferisco ad esempio, ai permessi "estesi" di un file oppure ai tempi di creazione/modifica.

TAR

Inizialmente sviluppato per scrivere dati sequenzialmente, tar è un utility per riunire in un unico blocco tanti file per distribuzione/archiviazione, preservanso allo stesso tempo informazioni sui permessi e sulla struttura delle cartelle. Tar è un ottimo modo per salvare file che si vuole recuperare a breve termine, per esempio se il sistema è a corto di spazio su disco, oppure è anche utile per spostare intere directory in un altro posto in quanto è in grado di preservare i permessi di accesso. Da notare che tar non è concepito per effettuare backup, l'ho incluso solo perchè potrebbe essere utile per piccole operazioni di salvataggio di file e cartelle

Esempio

```
tar -cvpzf /path/to/backups/backupxx-yy-zzzz.tar.gz --exclude=/home/somefile /home
```

Spiegazione

- c - crea un nuovo archivio.
- v - verbose mode, verrà scritto a schermo tutto il flusso delle operazioni eseguite da tar.
- p - preserva i permessi dei file.
- z - comprime i file usando l'algoritmo di gzip per rendere il tutto più compresso (richiede maggiore potenza di calcolo in compressione e decompressione).
- f - Specifica il path nel quale salvare il file, da notare che va specificato subito dopo.
- –exclude=/path - Seleziona i percorsi da escludere.

TAR VIA SSH

In questo modo è possibile avvalersi di ssh per trasferire la propria copia di backup al sicuro online, su un server via ssh `tar -cvpzf <...> | ssh user@host "( cat > ssh_backup01-03-14.tar.gz )"`

Pro

- Permette di salvare tutti i file in un unico blocco...
- ...ed eventualmente comprimerlo così da risparmiare spazio
- Rispetta la struttura delle cartelle
- Salva i file con i loro permessi originali

Contro

- Non conserva informazioni sul filesystem di partenza
- Comprimere una grande quantità di informazioni con un algoritmo avanzato può richiedere molta CPU

Rsync

Rsync è un utility che permette tramite rete, la sincronizzazione di file da una location ad un altra, eventualmente minimizzando i trasferimenti tramite algoritmi di compressione dati e delta encoding. Per determinare quali file vanno aggiornati rsync si basa sui confronti delle date di ultimo accesso al file, di conseguenza si avranno problemi qualora venissero fatte modifiche a file che non lansciano una traccia evidente (si può ovviare costringendo rsync a confrontare i file tramite checksum). Inoltre rsync non ha un sistema di gestione del backup avanzato quindi si limita esclusivamente alla sincronizzazione del file. Altra pecca di rsync è che la copia dei file viene salvata in chiaro nella destinazione.

Esempio

```
sudo rsync -aAXHEvz --exclude={"/dev","/proc","/sys","/tmp","/run","/mnt","/media","/lost+found"}
--delete --force --numeric-ids -e "ssh -i /home/r4yan2/.ssh/id_rsa" / r4yan2@grazio.so:/var/backups/rsy
```

Spiegazione

- -a, archive mode, abilita il mantenimento di tutti i metadati relativi ai file, come utente e gruppo di appartenenza, permessi, tempi di accessi e modifica, e abilita la discesa ricorsiva nelle cartelle
- -A, preserva gli attributi di tipo ACl
- -X, preserva gli attributi estesi
- -z, comprime i file durante i trasferimenti, utile se si vuole usare rsync via rete
- -H, preserva gli hard link
- -E, preserva la flag eseguibile
- –delete, cancella i file estranei nella destinazione
- –force, legato a –delete elimina anche le cartelle non vuote
- -e, permette di specificare la shell remota da utilizzare
- –numeric-ids, utile per evitare certi tipi di conflitti

Dealing with fat32

In caso si debba avere a che fare con device formattati con fat32 il comando si semplifica parecchio `rsync -rtv <src> <dst>` Questo avviene perchè fat32 non supporta molti dei metadati comunemente utilizzati sotto Linux quindi molti parametri di rsync possono essere evitati, infatti vengono attivati solo:

- -r, ricorsione nelle directory
- -t, preserva i tempi di modifica
- -v, incrementa l'output di rsync

Duplicity

Duplicity risolve ciò che rsync lascia in sospeso. Essendo un tool più voltato al backup dei file, propone un backup di tipo incrementale cifrato: all'avvio duplicity richiede di immettere una password per tenere al sicuro i dati o come alternativa si può passare come parametro una chiave gpg da usare per cifrare i dati. Tra le altre funzioni permette anche ti utilizzare un gran numero di protocolli/servizi come i classici ftp (file transfer protocol) o storage cloud (Amazon S3, Google Drive, ecc)

Esempio

```
duplicity /home/me sftp://uid@other.host//some_dir
```

- /home/me - percorso da salvare
- sftp://uid@other.host//some_dir - destinazione sottoforma di indirizzo sftp.

Nota Duplicity necessita che almeno la destinazione del backup sia sottoforma di url. Nel caso in cui sia locale deve essere scritto nella forma `file://path/to/something`

La prima esecuzione di duplicity ha lo scopo di creare il primo full backup che verrà salvato sulla destinazione. Esecuzioni successive dello stesso comando porteranno duplicity a controllare quali file sono stati cambiati dall'ultima esecuzione e provvederà a creare i blocchi "incrementali". Da notare che lo spazio richiesto con questo sistema dipende fortemente da quanto si modifica in proprio sistema: duplicity, infatti, è in grado di localizzare e salvare solo le modifiche.

Recupero

```
duplicity restore sftp://uid@other.host//some_dir /home/me
```

è il comando da utilizzare per recuperare i dati salvati, e dato che il backup è di tipo differenziale è possibile ripristinare i dati salvati ad una certa data:

`duplicity -t 3D restore sftp://uid@other.host//come_dir /home/me` xD - specifica la necessità di recuperare dei dati di x giorni prima che, ad esempio, sono stati cancellati per errore e poi è stato lanciato il comando di duplicity per avviare il backup

Infine è possibile eliminare dati che oramai non servono più se avessimo bisogno di spazio per nuovi backup:

`duplicity remove-older-than 30D sftp://uid@other.host//some_dir`

Pro

- Molto semplice da utilizzare
- Permette il backup differenziale
- Gran numero di funzioni per la gestione dei dati salvati
- Cifratura dei dati

Contro

- Non permette di gestire più di un host alla volta

dd & dd_rescue

dd è un utility di copia e conversione dati. Se non altrimenti specificato dd si limita a copiare dati da una sorgente alla destinazione. Uno dei principali utilizzi di dd è di copiare l'intero filesystem, comunque una scelta migliore potrebbe essere quella di usare mkfs sul filesystem di destinazione e usare dump/restore.

`dd if=/dev/hdx of=/dev/hdy conv=fdatasync`

- if - rappresenta il percoso da cui si legge
- of - il percorso in cui si scrive
- conv - indica dei parametri opzionali per la gestione del trasferimento dati, 'fdatasync' si assicura che tutti i dati vengano trasferiti prima che il comando termini l'esecuzione senza lasciare del lavoro incompiuto (nei buffer di sistema)

Quando si usa dd è importante assicurarsi che si opera sulle giuste partizioni di sistema poichè questo comando è potenzialmente distruttivo. Per controllare che le partizioni siano quelle corrette ci si può servire di tools come fdisk o gparted.

Pro

- Fornisce una copia perfetta della sorgente
- Mantiene inalterata la struttura del filesystem
- Permette di mantenere le informazioni relative alla tavola delle partizioni

Contro

- Richiede un drive di destinazione grande almeno quanto la sorgente
- Potenzialmente distruttivo
- Andrebbe usato come root per evitare che la copia fallisca per mancanza di permessi il che contribuisce ad aumentare la sua pericolosità

Restore dei singoli file/partizioni

A partire dall'immagine del disco si può accedere alle singole partizioni usando

`sudo losetup -Pf  /path/to/disk/image`

Questo comando mappa le partizioni presenti nell' immagine del disco sul primo loop device libero, che poi possono essere montati manualmente (saranno su /dev/loopXpY) o dal file manager. Questo per evitare di dover trovare un disco fisico dove scrivere tutta l'immagine, che spesso non è fattibile.

ddrescue

ddrescue è un tool simile a dd utilizzato quando si vuole recuperare il più possibile da un disco che sta per fallire. Infatti questo tool ignora gli errori presenti nel disco al contratio di molti altri che in presenza di errori interrompono le operazioni, così che si possa salvare il salvabile in casi estremi.

Dump & Restore

Dump e Restore sono altri comandi per creare e recuperare backup. Potrebbero non essere inclusi in tutte le distribuzioni quindi c'è bisogno di installarli esplicitamente. Dump opera creando una lista di file che sono stati modificati dall'ultimo dump e quindi pacchettizza i suddetti file in un largo blocco da archiviare in un device , o in alternativa è possibile optare per un sistema incrementale, rispetto a tar è più sofisticato in quanto è pensato ed ottimizzato per il backup, anche se è bene sapere che dump non è disponibile per tutti i filesystem, per esempio ReiserFS o FAT non sono supportati da dump al contrario di ext3 e ext4. Quello che in pratica fa dump è controllare la partizione, acquisirne il filesystem così da potersi muovere con una maggiore efficienza tra i file, ma ciò implica che è necessario effettuare il dump su ogni filesystem singolarmente, e inoltre dump permette solo il backup dei dati che sono in locale, non permette ad esempio backup di dati presenti su di un NFS Restore dal suo lato ha molte opzioni, tra cui le più importanti:

- -i per il restore interattivo di file indivibuali e/o directories
- -r per il recupero completo di un intero filesystem
- -x che richiede in recupero non interattivo di un dato file

Pro

- In circolazione da ormai molto tempo
- Ottimizzato per il backup
- Lavora a livello di filesystem
- Permette backup di tipo incrementale e differenziale
- Restore possiede opsioni specifiche per ogni esigenza

Contro

- Lavora solo su percorsi locali
- Non disponibile per tutti i filsystem

rdiff-backup

Esempi

```
rdiff-backup dir1 user@system::/dir2
```

```
rdiff-backup dir1 dir2
```

rdiff-backup è un utility in circolazione ormai da molto tempo che permette il backup in locale o anche in remoto utilizzando il metodo incrementale. Rdiff-backup è un utility appositamente studiata per il backup dei dati e perciò troviamo funzioni avanzate come la replicazione della struttura delle sottocartelle, permessi, dev files, e inoltre rdiff-backup può operare in modo efficiente rispetto alla banda, utilizzando apposite funzioni di pipe.

Pro

- preserva la struttura delle cartelle, links, permessi, attributi estesi, meta- dati.
- ottimizzato per la rete
- facile da utilizzare
- permette backup incrementale
- mantiene dei log

Riferimenti & bibliografia

- http://en.wikipedia.org/
- https://help.ubuntu.com/community/BackupYourSystem/TAR
- https://wiki.archlinux.org/
- Giornalinux 2.0 N14 - www.poul.org
- http://duplicity.nongnu.org/
- http://www.nongnu.org/rdiff-backup/index.html
- Unix and Linux system administration handbook, fourth edition, EviNemeth, Garth Snyder, Trent R. Hein, Ben Whaley