Backup (v1.0)

Andrea Grazioso

27 Marzo 2014

Indice

Capitolo 1

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 molto 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 di un device diverso da quello su cui si opera normalmente cosicché sia possibile un eventuale ripristino del sistema o di singoli file in caso di perdita di questi ultimi o in casi più sfortunati, un recupero totale del sistema in seguito al malfunzionamento di un hard disk che in caso di applicazioni con uso intensivo di dati come il campo audio è un eventualità tutt'altro che remota, il furto del PC, la 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, ma è opportuno puntualizzare che utenti diversi avranno esigenze di backup diverse: ad esempio l'home user non è obbligato ad effettuare il backup ma in ogni caso è un operazione caldamente consigliata, diventa una necessità per chi sul proprio pc conserva dati sensibili, progetti e documenti importanti la cui perdita potrebbe causare notevoli danni a livello personale, e infine per chi lavora utilizzando un computer potrebbe addirittura essere imposto dalle politiche aziendali, di conservare, o di affidare a terzi la copia del proprio sistema.

Analogamente per gestori o sistemisti di VPS/server dedicati/datacenter è fondamentale conservare copie multiple differenziate 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 per effettuare il 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 unità flash esauriscono la loro vita dopo approssimativamente 100'000 cicli di scrittura,

invece gli hard disk vanno incontro a problemi molto diversi come 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, ad esempio è inutle conservare il drive di backup nella stessa borsa con cui si trasporta il pc, infatti qualora vi venisse sottratta la borsa perdereste sia il pc che il drive di backup. In caso si potrebbe anche optare per il backup via rete, possedendo un NAS (Network Attached Storage) domestico oppure via SSH i dati possono essere spostati su di un server, o ancora via NFS (Network File System) su di una macchina remota.

Le possibilità sono veramente tante e sono tutte aperte a possibili combinazioni dettate dalla fantasia individuale, è importante ricordare però che in caso di crash del sistema il backup dovrebbe essere facilmente accessibile per il ripristino, e inoltre il backup rappresenta sempre una copia del proprio sistema quindi mettere quei dati in rete privi di una qualsiasi forma di crittografia vi espone a dei potenziali rischi.

Capitolo 2

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, ed in questa sezione andrò a spiegare meglio quali tipi di supporti sono più indicati per quali operazioni.

2.1 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...
- ...che sono molto ingombranti...
- ...e costose

2.2 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

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.

2.3 Dispositivi Ottici

CD, DVD e Blu-ray rappresentano un buon supporto per il backup. Essi 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, 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, la cui velocità è variabile a seconda del costruttore e modificabile dall'utente a seconda del grado di sicurezza con il quale si vuole salvare i dati. Degna di nota la loro impermeabilità e la resistenza al manetismo, invece per quanto riguarda i danni fisici, resistono anche bene ai danni superficiali lievi grazie a moduli di controllo degli errori, ma in caso di danno profondo si rischia di perdere i dati.

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 stati manipolati in modo malevolo o eliminati dal drive. Invece altri formati di CD, DVD e Blu-ray sono riscrivibili, ma questa caratteristica e limitata a poche centinaia di riscritture.

Per chi pensa sia assurdo che al giorno d'oggi ci si affidi a questi supporti per il backup riporto il caso di Facebook, l'ormai noto social network, che ha intrapreso una campagna di salvataggio del proprio archivio dati utilizzando Blu-ray, che secondo il famoso social network saranno impiegati per contenere svariati Petabyte di informazioni, foto, video di "cold storage", cioè dati poco utilizzati o addirittura inutili, che però di devono impegnare a conservare per vari motivi (legali, amministrativi, ecc.).

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 vibrazioni e movimenti

Schematizzando

	Dim Max	Lifetime	Min Write Speed	Max Write 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

2.4 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 il 2.0 60 MB/s e 3.0 625 MB/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.

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 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

USB	Velocità Dichiarata [MB/s]	Velocità Reale [MB/s]	rilasciato	lunghezza max cavo [m]
1.0	1.5*	-	1996	3
2.0	60	35	2000	5
3.0	625	200	2010	3**

(*) In realtà lo standard USB 1.0 è stato più volte ridefinito, infatti era inizialmente pensato per il collegamento di periferiche di input come mouse e tastiera che sono per definizione a bassissimo bitrate, la velocità iniziale delle prime porte era di circa 8 volte più bassa delle ultime.

(**) Questo valore non è mai stato definito nello standard USB3.0 pertanto rimane solo "consigliato" e non è da considerarsi un vincolo reale.

Nota nel 1998 è stato rilasciato lo standard USB 1.1 che è stato impiegato in alcune piattaforme di uso comune (ad esempio le play station 2 slim erano equipaggiate con questo genere di porte), si noti che anche con questo ulteriore upgrade le porte USB non supereranno la velocità dei DVD fino all'introduzione dello standard 2.0

2.5 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 e della buona velocità di trasferimento dati, sopratutto grazie alle porte (SATA/IDE/PATA) con il quale vengono connessi internamente alla macchina. 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 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.

Le note a loro sfavore sono la loro bassa resistenza a urti e lesioni in generale, infatti essendo degli ibridi tra tecnologia meccanica e magnetica sono molto sensibli agli urti e agli spostamenti bruschi che potrebbero danneggiare più o meno gravemente il drive.

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

2.6 SSD

Questo tipo di memorie è relativamente nuovo: la loro origine 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, e inoltre cioè permette di avere delle velocità irraggiungibili con qualsiasi altro drive (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 li rende più soggetti a guasti degli HDD, perciò questi ultimi sono ancora da preferire nel caso si debba conservare una grande quantità di informazioni.

Nel caso di VPS/server dedicati essi vengono utilizzati principalmente per assicurarsi le massime prestazioni dalla machcina 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 o 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

2.7 Backup Online

Piùttosto che occuparsi personalmente del device su cui effettuare il backup si potrebbe optare per affidarsi a servizi di archiviazione/backup di dati online come ad esempio Dropbox, Amazon Glacier, OVH Cloud Storage. Tenere i backup in remoto aiuto 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, altrimenti è bene assicurarsi personalmente della loro protezione, ad esempio creando un archivio protetto da password.

Inoltre questi servizi di archiviazione online offrono, in genere, un certo grado di scalabilità, cioè la possibilità di acquistare spazio aggiuntivo che andrà quindi ad ampliare la pool di spazio già 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.

Dropbox Nato come metodo di file sharing più che di backup il servizio di Dropbox consente di sincronizzare file in una apposita directory nel proprio PC accessibile da web, smartphone e altri PC connessi con lo stesso account Dropbox. Il servizio offre anche la possibilità di riprendere file eliminati dalla cartella condivisa e mette a disposizione 2GB gratuiti per tutti i nuovi utenti, ma occhio alla sicurezza! Dropbox infatti è nota per avere delle politiche molto lascive sulla sicurezza dei dati, infatti i TOS del servizio non danno garanzie effettive sulla privacy degli utenti registrati e quindi qualora si decida di utilizzare questo servizio è bene utilizzare gli accorgimenti proposti precedentemente.

Amazon Glacier Servizio di storage low-cost di Amazon, pensato per offrire servizio di backup dei dati, Glacier offre servizi di trasferimento sicuro dei dati tramite SSL e crittografa automaticamente i dati usando chiavi simmetriche AES a 256-bit. Permette di scalare il proprio piano tariffario adattandosi alle esigenze dell'utilizzatore mantenendo inalterati tutti i servizi accessori. E' un ottimo supporto da utilizzare per il "cold storage", cioè il deposito di file che non si intende utilizzare a breve termine.

In quanto a costi Glacier offre banda in upload pressapoco illimitata, mentre per effettuare il download dei dati salvati potrebbe essere richiesto un accredito, oltre al pagamento dello storage.

OVH Public Cloud Storage Analogo di Glacier offerto da OVH questo sistema di cloud storage è un altra ottima opsione per l'archiviazione dei propri dati in quanto effettua la ridondanza dei dati, assicura una reeribilità dei dati salvati al 99.99%, la risondanza dei dati in 3 datacenter, permette download e upload sicuri utilizzando il protocollo HTTPS ed è a suo modo scalabile.

BackBlaze Questo singolare servizio vi offre la possibilità di archiviare una quantità di dati praticamente illimilata per soli $5/mese

Capitolo 3

Metodi di Backup

Quì entra in gioco una prima distinzione fondamentale, quella tra Snapshot e Backup, il primo più capiente 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, per ottimizzare spazio utilizzato, versioning dei file, recupero selettivo dei file, ecc.

A seconda di ciò che si deve mettere al sicuro ci possono adottare più approcci al backup:

3.1 Snapshot

Cioè l'immagine di sistema, una copia a tutti gli effetti del proprio sistema, file e metadati compresi, pesante quanto il sistema di partenza necessita un drive grande a sufficienza ed eventualmente di grandezza multipla se si prevede di usare quello stesso drive per backup distanziati nel tempo. Noto anche come "raw partition backup", "disk image", "dump", o più comunemente "snapshot", lo scopo di questa tecnica è di copiare settore per settore, bit per bit un volume o un disco intero preservando l'esatta struttura del volume di partenza copiando in caso di hard disk interi anche settori di boot, tavola delle partizioni e struttura delle partizioni.

Tool di utilizzo frequente per questo scopo sono:

- dd
- rsnapshot

3.2 Backup

Come accennato in precedenza, questo è un approccio mirato a singoli file o interi filesystem. E' una tecnica che abbiamo sperimentato un po' tutti inconsapevolmente copiando dei file su una periferica esterna, per renderli più

facilemente trasferibili o accessibili. Questo metodo sicuramente non ricrea la complessità di un sistema operativo per cui non avremmo possiblità in caso di un crash di sistema, in ogni caso da questo punto di partenza andiamo alla scoperta di tecniche avanzate per la gestione e l'ottimizzazione delle nostre copie di Backup.

Tra i tool più basilari per la manutenzione dei backup troviamo:

- duplicity
- rdiff-backup

3.3 Metodi Avanzati

Tra le tecniche avanzate che vedremo adottate da vari programmi specifici per il backup troviamo la possibilità di trasferire file in directory remote, tenere indici dei file trasferiti, riprendere trasferimenti interrotti, e gestire le versioni dei file come vediamo sotto in questi due differenti approcci:

- Incrementale, aggiunge all'ultimo backup, il materiale ha subito dei cambiamenti dall'ultimo backup o incremento. Più sicuro di quello differenziale ma necessità di attenzione nel ripristino.
- 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.

Capitolo 4

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 come strutturare i dati da salvare, ma c'è un ultimo step da rivedere prima di mettere effetivamente mano ai tools di backup.

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 oppure fare un dump di /home.

In caso di backup completo del sistema 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 ed eliminare eventuali file temporanei lasciati in giro per il filesystem.

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.

Inoltre va tenuto in considerazione che i percorsi /tmp e /media contengono rispettivamente file temporanei e percorsi di mount di periferiche esterne quindi non devono essere inclusi nel processo di backup, così come le directory /sys e /proc. Analogamente /var ed /etc contengono principalmente file di configurazione, per cui cambiano raramente in un utilizzo domestico, e si potrebbe anche decidere di soprassedere se non si è effettuata personalizzazione del sistema successiva all'installazione. D'altra parte se volete fare il backup del vostro server/VPS è scontata la necessità di salvare i file di configurazione ad esempio del webserver (Apache2, Nginx, Lighttpd) o di PHP. A questo proposito vedremo come sia possibile ottenere tramite semplici comandi da terminale una copia esatta di un intero sistema.

Infine per la perfetta riuscita di un backup sarebbe ottimale ridurre al minimo l'attività nel filesystem, e il modo migliore per farlo è avviare il sistema

via livemedia; ovviamente come operazione è un po' scomoda e su questo potremmo anche chiudere un occhio.

Capitolo 5

Programmi

Quì si apre veramente un molto di scelte e possibilità, sostanzialmente dettate dalle proprie esigienze, ci sono 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 a specifici flag abilitati nel mount del filesystem per specifiche funzioni come l'impossibilità di eliminare file.

Passeremo ora ad esaminare i programmi che sono maggiormente utilizzati per il backup del sistema

5.1 TAR

Inizialmente sviluppato per scrivere dati sequenzialmente (per effettuare backup sui nastri magnetici), tar è ora 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 segue i simlink di default ma c'è un opzione apposita per farlo

```
tar -cvpzf /media/BackupStorage/backup01-03-14.tar.gz
    --exclude=/backup.tar.gz /
```

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 assoluto nel quale salvare il file, nell'esempio lo salva in un disco esterno, come potete vedere dall'esempio è buona norma ordinare i backup conservando anche la data in cui sono stati eseguiti così da poterli catalogare più facilemnte in seguito.

/media/BackupStorage/backup01-03-14.tar.gz - Specifica il path in cui salvare il backup, nella fattispecie viene salvato nell'unità esterna chiamata "BackupStorage"

–exclude=/path - Seleziona i percorsi da escludere, quindi /tmp /dev /media/<altre periferiche>

/ - a partire da /root viene fatta ricorsivamente l'archiviazione dei file in tutte le cartelle sottostanti.

5.1.1 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 –cvpz <...> / | ssh user@backuphost
    "( cat > ssh_backup01-03-14.tar.gz )"
```

| - simbolo di pipe indica che l'output del comando a sinistra verrà reindirizzato tramite quello che segue

ssh - Secure Shell Host, tramite il quale ci si può collegare a "backuphost" come "user"

cat - utility basilare per effettuare lettura/scrittura di/su file

> - altro simbolo usato per indicare il reindirizzamento dell'output

ssh-backup01-03-14.tar.gz - rappresenta il file che verrà salvato sulla destinazione remota, in questo caso verrà salvato nella home di "user"

Pro

- Permette di salvare tutti i file in un unico blocco...
- ...ed eventualmente comprimerlo così da risparmiare spazio
- Rispetta la struttura delle cartelle
- Non segue i simlink*
- 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

Note (*)Tar NON segue i simlink, e ciò può essere un vantaggio come uno svantaggio in quanto permette di includere filesystem come /proc e /sys senza preoccupazioni, ma in caso si voglia volutamente che tar segua i simlink bisognerà specificarlo con un opzione apposita.

5.2 Rsync e Duplicity

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. Duplicity risolve questo problema, essendo un tool più voltato al backup dei file, propone un backup di tipo differenziale cifrato, quindi all'avvio duplicity richiede di immettere una password per tenere al sicuro i dati o come alternativa si può esare una chiave publica per cifrare i dati da decifrare con la propria chiave privata, e tra le altre funzioni permette anche ti utilizzare un gran numero di protocolli/servizi come i classici ftp (file transfer protocol) o strumenti cloud (Amazon S3, Google Drive, ecc)

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

/home/me - percorso da includere nel backup con tutte le relative sottocartelle

sftp://uid@other.host//some_dir - destinazione sottoforma di indirizzo sftp, nello specifico ci si collega ad "other.host" come user "uid", e il backup avrà posto in "some_dir"

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 "differenziali". 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 le modifiche e salvare esclusivamente i blocchi del file che hanno subito delle modifiche.

```
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

- Sostanzialmente nessuna, è ottimizzato per il backup di una singola unità più che un insieme di macchine.

5.3 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
```

if - rappresenta il percoso da cui si legge

of - il percorso in cui si scrive

quando si usa dd o l'analogo dd_rescue 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

Note dd_rescue non è una vera e propria utility di backup ma è indicata per il recupero di dati da supporti sul punto di fallire definitivamente, è quindi indicato per assolvere tali funzioni in quanto possiede delle procedure ottimizzate per queste funzioni.

dd via ssh

```
dd if=/dev/sda | ssh username@servername.net
    "dd of=/file"
```

| - pipe, indica che l'output del comando a sinistra verrà reindirizzato al seguente

ssh - Secure Shell, tramite il quale ci si collega a "backuphost" come "user"

"" - comando (o parte) da eseguire come remote host

Nota Per utilizzare questa funzione è importante ricordarsi di effettuare un corretto setting dei permessi.

5.4 Dump & Restore

Dump e Restore sono comandi per creare e recuperare backup. Questi programmi circolano oramai da molto tempo, sono stati ampiamente testati ed hanno dei bug che è bene conoscere prima di cimentarsi nel loro utilizzo. 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 opsioni, tra cui le più importanti sono sicuramente:

-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 esigienza

Contro

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

5.5 rdiff-backup

```
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.

Come detto in precedenza è possibile utilizzare rdiff-backup e ssh per effettuare il backup di dati su una directory remota e solo le differenze verranno trasmesse alla destinazione.

Pro

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

5.6 Bacula e Amanda

(work in progress...)

Capitolo 6

Riferimenti & bibliografia

- http://en.wikipedia.org/
- https://help.ubuntu.com/community/BackupYourSystem/TAR
- https://wiki.archlinux.org/
- http://www.tomshardware.com/reviews/usb-3-uas-turbo,3215-2.html
- http://data-backup-software-review.toptenreviews.com/removable-backup-media.html
- http://www.tomshardware.com/reviews/usb-3-uas-turbo,3215-2.html
- Giornalinux 2.0 N14 - www.poul.org
- http://duplicity.nongnu.org/
- http://aws.amazon.com
- http://Dropbox.com
- http://www.backblaze.com/
- http://www.nongnu.org/rdiff-backup/index.html
- http://amanda.org/
- http://www.bacula.org/en/
- Unix and Linux system administration handbook, fourth edition, Evi Nemeth, Garth Snyder, Trent R. Hein, Ben Whaley