

LinuxDay 2009 @ ERLUG

*ALLA RICERCA DEL DATO PERDUTO:
UN ANNO CON BACKUPPC*

*PRESENTAZIONE A CURA DI
MASSIMILIANO MASSERELLI*

Alla ricerca del dato perduto

Concetti generali

- ◊ Strategie di backup
- ◊ I grandi nemici
- ◊ I preziosi alleati

Nella pancia di BackupPC

- ◊ Installazione
- ◊ Dove stanno i dati?
- ◊ Automazione

Quello che le docu non dicono

- ◊ Considerazioni di sicurezza
- ◊ Dove mettere il repository
- ◊ Spostare il repository
- ◊ Backup off-site

Strategie di backup: backup religioso

Il sistema più diffuso nel mondo: a scelta, si prega perchè non si perda nulla o si insultano diversi dei più o meno a caso quando inevitabilmente si rompe qualcosa.

- ✓ Economico
- ✓ Veloce
- ✓ Non richiede competenze particolari
- ✓ Nessuna manutenzione



Strategie di backup: on-line/off-line

On-line: i dati sono salvati su supporti non rimovibili

- ✓ Nessuno spreco di spazio
- ✓ Accesso immediato ai dati
- ✓ Salvataggio in parallelo
- ✓ Controllo automatico dell'integrità dei dati
- x In caso di guasto all'hw, i dati possono essere compromessi
- x I dati sono conservati insieme all'hw

Off-line: i dati sono salvati su supporti rimovibili (nastri, CD, ecc...)

- ✓ Duplicazione dei dati su più supporti
- ✓ Indipendenza del supporto dall'hw
- x Salvataggio seriale
- x Possibilità di mantenere i supporti lontano dall'hw

Strategie di backup: storicizzazione

Oltre agli eventi disastrosi un backup deve poter proteggere anche da imprevisti meno evidenti:

- ✓ Cancellazione accidentale
- ✓ Corruzione di file
- ✓ Compromissione di una macchina

Mantenere copie di backup ad intervalli regolari:

- In caso di supporti rimovibili, alcuni supporti vanno dedicati all'archiviazione storica
- In caso di backup on-line, multiple copie dello stesso file vengono memorizzate sullo stesso supporto

I grandi nemici: Murphy

Se qualcosa può andare storto lo farà!

- ◊ I dati memorizzati sono leggibili?
- ◊ I dati memorizzati sono quelli che ci si aspetta?
- ◊ Dal backup è possibile effettuare un restore “bare metal”?
- ◊ Quali sono i tempi di restore per un singolo file/un'intera macchina?
- ◊ Quante e quali persone sono capaci di effettuare un restore?



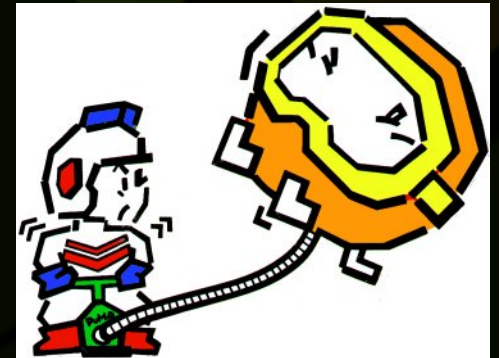
I grandi nemici: Moore

Prima legge di Moore (Gordon Moore, cofondatore di Intel con Robert Noyce – ca. 1970):

- «Le prestazioni dei processori, ed il numero di transistor ad essi relativo, raddoppiano ogni 18 mesi.»

Questa “legge” applica a molte cose collegate all'informatica, compresi i supporti di massa

- I sistemi di backup devono essere scalabili
- Si deve prevedere per tempo la necessità di incrementare lo spazio disponibile sui supporti utilizzati per l'archiviazione
- La necessità di spazio non aumenta linearmente con il numero di computer connessi alla rete



I grandi nemici: MS Windows

- Per ragioni storiche collegate alle origini “single user” di MS-Windows, le applicazioni aprono i file in uso esclusivo, impedendone la lettura e quindi il backup.
 - Registry e sua copia di backup
 - File utilizzati per memorizzare la posta (.pst)
 - Buona parte dei file utilizzati per le preferenze degli utenti (\users)
 - File di swap
- Non basta una semplice copia del file-system per trasferire il S.O. Installato su un'altra macchina
 - Durante l'installazione vengono “scelti” ed installati driver fondamentali per il caricamento del S.O.
 - Il codice di attivazione della copia di Windows è legato all'hardware della macchina

I grandi nemici: DBMS

- Praticamente tutti i sistemi DBMS (Mysql, Oracle, PostgreSQL, ecc...) salvano i loro dati in maniera “complessa”
 - I dati non vengono salvati in maniera atomica
 - Una semplice copia del filesystem non garantisce che il risultato sia consistente
 - Una base dati inconsistente può essere male

I preziosi alleati: E-SATA

Nella necessità di trasportare grandi volumi di dati, storicamente ci si rivolgeva a supporti magnetici su nastro. L'alternativa oggi è usare hard-disk con supporto E-SATA.

- ◊ Standard
- ◊ Basso costo interfaccia
- ◊ Buon rapporto prezzo/capacità
- ◊ Ampia capacità
- ◊ Rimuovibile
- ◊ Hot plug
- ◊ Buona velocità di trasferimento
- ◊ Nessun carico aggiuntivo per il processore

I preziosi alleati: VSS/LVM

VSS in ambito Microsoft ed LVM in ambito Linux permettono di creare degli “snapshot” dei dispositivi di massa, in modo trasparente e senza interrompere l'operatività.

- ◊ Garanzia di immutabilità dello snapshot
- ◊ Utilizzo tecnologia copy-on-write
- ◊ Al termine della sua vita lo snapshot viene distrutto, senza lasciare tracce

BackupPC: installazione

```
# aptitude install backuppc  
# htpasswd /etc/backuppc/htpasswd backuppc  
# firefox http://localhost/backuppc
```

Davvero, non serve niente più di questo!

Ok, il resto della configurazione è web based, ed ogni voce è documentata on-line

BackupPC: configurazione

Dicevamo che tutta la configurazione è web based, ma per chi non ha modo di usare un browser?

- ◊ `/etc/backuppc/config.pl` (configurazione generale)
- ◊ `/etc/backuppc/localhost.pl` (configurazione di un host)

I file di configurazione sono in linguaggio PERL. Per la maggior parte, si tratta solamente di definire delle variabili e delle strutture che descrivono il comportamento da tenere per il backup ed il modo per raggiungere i dati

Morale: la configurazione web se la regna.

BackupPC: test

All'installazione, viene creato un host predefinito (localhost) che tramite tar esegue il backup della directory /etc della macchina su cui è stato installato il software

- ◊ Eseguire un backup manuale “full”
- ◊ Eseguire un backup manuale “incrementale”
- ◊ Verificare cosa è stato salvato
- ◊ Provare un restore dei dati salvati (archivio compresso o restore “in place”)
- ◊ Ripetere i punti precedenti una dozzina di volte (Murphy)

BackupPC: il repository

- ◊ BackupPC salva i dati dei backup delle diverse macchine in `/var/lib/backuppc/pc/nomemacchina/numeroprogressivo`
- ◊ I file “veri” vengono salvati in `/var/lib/backuppc/cpool/...` (file compressi) o in `/var/lib/backuppc/pool/...` (file non compressi)
- ◊ Il formato è “quasi standard”: il nome del file è un hash, e nel caso di file compressi, nell'header sono salvate alcune informazioni utilizzate dal software per la gestione
- ◊ I file salvati per ogni singola macchina vengono salvati in pool, e quindi linkati (hardlink) in `/var/lib/backuppc/pc/nomemacchina/numeroprogressivo`

BackupPC: considerazioni di sicurezza

- ◊ Permettere ad una macchina di fare un backup automatico significa che la sicurezza dei dati è pari alla sicurezza della macchina che esegue il backup.
- ◊ Bisogna scegliere un “trasporto” adeguato: backupPC offre diverse possibilità (tar, rsync, samba) in base alla struttura della rete in esame
- ◊ Il metodo più diffuso attualmente (per il backup di macchine linux) sembra essere rsync lanciato tramite ssh autenticando su chiave crittografica

BackupPC: rsync+ssh

- ◊ Configurare un utente (backuppc) sulla macchina remota con password
- ◊ Dalla macchina che ospita backuppc, copiare la chiave RSA dell'utente backuppc

```
$ ssh-copy-id nomeoipdellamacchinaremota
```

- ◊ Disabilitare la password dell'utente backuppc sulla macchina remota
- ◊ Verificare che sia possibile connettersi tramite ssh

```
$ ssh nomeoipdellamacchinaremota
```

- ◊ Configurare opportunamente sudo sulla macchina remota, in modo che backuppc possa eseguire rsync in modalità server con privilegi di root modificando /etc/sudoers:

```
backuppc ALL = NOPASSWD: /usr/bin/rsync ...
```

- ◊ Configurare backupPC per eseguire rsync sulla macchina remota passando per sudo:

```
$sshPath -q -x -l backup $host sudo $rsyncPath $argList+
```

BackupPC: filesystem per il repository

- ◊ Pool e directory delle macchine conterranno un grande numero di file, per la maggior parte hardlink
- ◊ Grande consumo di inode rispetto alle dimensioni medie dei file
- ◊ Il consenso sembra essere per l'utilizzo di xfs, che non ha limiti decisi al momento della creazione sul numero massimo di inode e sembra essere performante nella gestione di molti file piccoli

BackupPC: spostare il repository

- ◊ Ripetete con me:

IO NON SPOSTERÒ IL REPOSITORY

- ◊ La copia di un repository intero, pool e directory di backup, per dimensioni contenute (attorno ai 300Gb) può richiedere giorni di attesa
- ◊ Se proprio c'è bisogno, si può usare “cp -a” o “rsync -avH”
- ◊ Se il repository si trova su un disco fisico, il miglior approccio è usare dd (copia fisica) e quindi estendere il filesystem una volta copiato (l'implementazione è lasciata come esercizio per lo studente)

BackupPC: backup offsite

- ◊ Duplicare il repository off-site può essere antipatico
- ◊ BackupPC mette a disposizione un trasporto virtuale, chiamato “archive”, che permette di creare un archivio compresso standard di un determinato backup
- ◊ Udev può essere nostro amico:
- ◊ `/etc/udev/rules.d/99-backupoffsite.rules`:
`ACTION=="add",`
`ENV{ID_FS_LABEL_ENC}=="OFFSTBCKP",`
`RUN:="/usr/local/sbin/backup.sh"`

BackupPC: esempio di script per il backup offsite

```
#!/bin/bash
/usr/bin/logger "backup offsite iniziato"
/bin/mount /media/backup
rm -rf /media/backup/*
/bin/umount /media/backup
/usr/bin/sudo -u backuppc
/usr/share/backuppc/bin/BackupPC_archiveStart special-
tapearchive backuppc localhost tblogger tbserver tbterminal
tecnodell tecnopass
/usr/bin/logger "backup offsite in corso"
```

Link utili

<http://backuppc.sourceforge.net/>

Sito ufficiale di BackupPC

<http://www.zmanda.com/>

Sito commerciale di supporto a strumenti di backup open source

<http://www.bacula.org/>

Sistema di backup open source

<http://www.amanda.org/>

Sistema di backup open source

ARRIVEDERCI E GRAZIE PER TUTTO IL PESCE!

QUESTA PRESENTAZIONE È STATA REALIZZATA CON
L'AUSILIO DI:

OPENOFFICE IMPRESS
THE GIMP
OPENCLIPART LIBRARY
GNU/LINUX OS

DOMANDE?