

LVM, SNAPSHOT, RSYNC, BACULA: Backup professionale a costo (quasi) zero

Autori: D. Verzulli - Università "G. D'Annunzio"

Abstract

L'archiviazione ed il backup efficiente dei dati fondamentali attinenti l'intera infrastruttura di una Rete d'Ateneo, costituisce generalmente un problema particolarmente complesso.

L'ottimale raccolta ed archiviazione di dati "immutabili" (Es.: LOG) ed il backup di informazioni particolarmente dinamiche (mailbox, web), impongono requisiti sull'infrastruttura di backup differenziati e complessi, specie se valutati in termini di minimizzazione del costo di implementazione (ossia utilizzando esclusivamente software Open-Source e limitando gli investimenti in hardware).

La scelta effettuata dal NOC della "d'Annunzio" è costruita con:

- l'utilizzo di LVM su un totale di 1.5TB di storage (facilmente upgradabile a 4/6TB) locale ad un PC entry-level (1xP4-3.2Ghz/2GB);
- l'impiego di SNAPSHOT multipli (read/only) per mantenere "on-line" 7 versioni di backup di posta e web, minimizzando lo storage fisico richiesto;
- una (piccola) serie di (piccoli) script (PERL) che automatizzano l'RSYNC dei dati fra i server di produzione ed il backup-PC, gestendo opportunamente la creazione/distruzione degli SNAPSHOT giornalieri.

In tal modo diventa banale recuperare la pagina X oppure la mailbox Y purché la richiesta arrivi entro 7 giorni dalla data della modifica.

Per le esigenze di archiviazione a lungo termine (LOG in primis, ma anche POSTA e WEB), all'architettura sopraindicata è stata aggiunta una Tape-Library di nuova acquisizione (IBM TS/3100 da 22+1 nastri LTO4 da 800GB x ca. 8000 euro) collegata via SCSI ad uno dei server di produzione particolarmente "scarico" (1xP4-2.66Ghz/1.5GB/2xGeth) sul quale è stato installato BACULA (Open Source).

La configurazione di BACULA è stata effettuata in due fasi:

- configurazione per la gestione della tape-library (load/unload di nastri identificati attraverso un barcode e relativo lettore);
- configurazione delle dinamiche di backup (backup totali, incrementali, differenziali e relativi pool, retention, etc.).

I risultati sono andati oltre ogni aspettativa: BACULA si è mostrato un software decisamente maturo in termini di affidabilità, gestibilità, robustezza. Abbiamo effettuato anche dei tentativi (riusciti) di "recovery" ricostruendo il backup dal

9° WORKSHOP GARR
GARR – The Italian Academic & Research Network

singolo nastro (con una installazione from-scratch di OS+Bacula). Abbiamo anche testato sul campo l'effettivo "restore" di dati che un'utente ha richiesto, più vecchi di una settimana.

Il sistema è in produzione da settembre 2008 e, nella sua fase "a regime", ha richiesto (come previsto...) l'impiego di una LAN ad-hoc, con collegamento Giga-Ethernet fra il pc ed il server di backup (+/- 120Mbps sustained).

Attualmente stiamo valutando:

- l'uso del "client" BACULA "nativo" (linux _E_ windows), in alternativa all'rsync;
- l'ottimizzazione dei processi di "disaster recovery" (cdrom-based);
- l'ottimizzazione della reportistica.