

Matteo Castellini



Gestione dei volumi con LVM

Linux Day 2007



Cos'è LVM?

LVM è il Logical Volume Manager di Linux, il software che permette di fare logical volume management.

LVM fornisce un'astrazione della memoria di massa a più alto livello rispetto al tradizionale schema di dischi e partizioni, dando maggiore flessibilità nella gestione della stessa.



Bla bla bla storico

Il logical volume management non è cosa nuova nel campo dei sistemi operativi.

In Linux la prima implementazione (LVM1) risale al 1997.

Durante lo sviluppo del kernel 2.5 furono proposte due implementazioni concorrenti: EVMS (IBM) e LVM2.

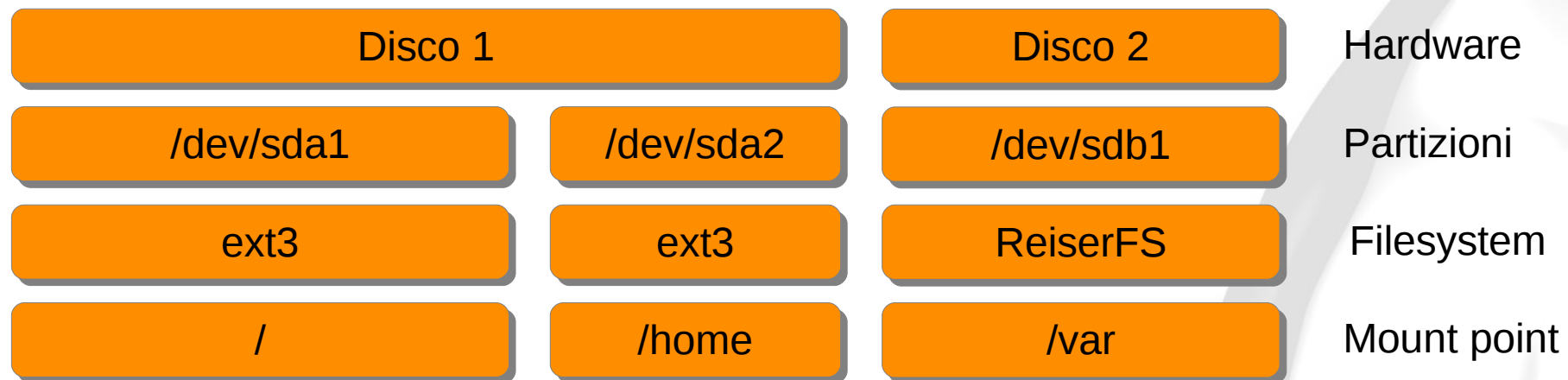
Alla fine la spuntò LVM2, EVMS rimase fuori dal kernel spostando il suo interesse sui tool in userspace.



Storage: the old way

La gestione dei volumi senza LVM è tanto semplice quanto scomoda perché è strettamente legata all'hardware.

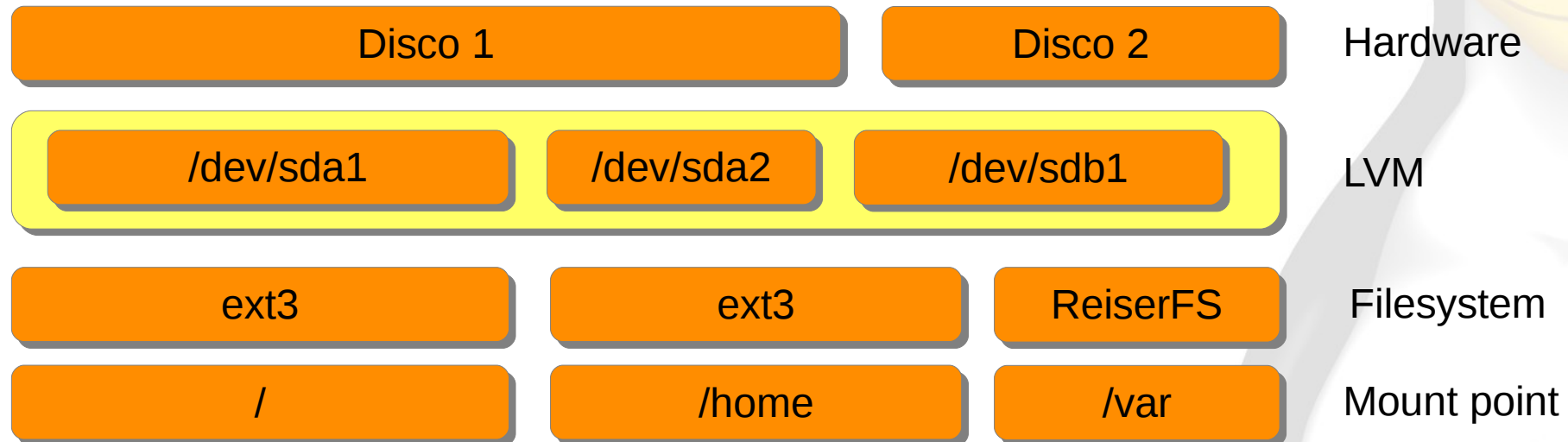
Due attori in scena: partizioni e filesystem.



Flessibile come una lastra di marmo. :)

Storage: the LVM way

La gestione dei volumi con LVM è più flessibile, per far questo si introduce uno strato di astrazione.



LVM si incarica di gestire la parte software a più basso livello per permettere di fare tante cose carine. :)

Le cose carine

- Gestire e nominare volumi logici che occupino uno o più volumi fisici.
- Creare, ridimensionare, spostare ed eliminare volumi logici anche “a caldo”.
- Fare striping di volumi logici (interi o parti) tipo RAID 0.
- Fare mirroring di volumi logici (interi o parti) tipo RAID 1.
- Creare snapshot di volumi logici in lettura e scrittura.

LVM è solitamente associato a grandi installazioni ma è ugualmente utile in piccole sistemi con uno o due dischi.



Nel piccolo... /1

Un problema comune a chi si avvicina a Linux: gestire il partizionamento su un disco “piccolo” (40 GB).

/boot	/dev/hda1	100 MB
swap	/dev/hda2	1 GB
/	/dev/hda3	10 GB
/home	/dev/hda4	28 GB

Per vari motivi (collezione mp3, editing multimediale, virtualizzazione) lo spazio allocato a /home è insufficiente. Rimane tuttavia diverso spazio inutilizzato su /.

Come fare a risolvere il problema?

Nel piccolo... /2

- Le opzioni a disposizione non sono le migliori:
 - Formattare, cambiare lo schema delle partizioni e reinstallare.
 - Comprare un nuovo disco.
 - Creare dei symlink che permettano di sfruttare lo spazio inutilizzato in /.

Potendo disporre di LVM la cosa potrebbe essere risolta in modo molto più semplice e indolore ridimensionando / e aggiungendo lo spazio ricavato a /home senza nemmeno dover smontare i filesystem.

Nel caso di aggiunta di un nuovo disco sarebbe inoltre possibile allargare /home trasparentemente.



... e nel grande

I benefici di LVM sono più ovvi su sistemi con molti dischi.

Gestire un grosso pool di dischi è dispendioso in termini di tempo ed è complesso gestirne un insieme eterogeneo.

Gestire le esigenze di spazio dei singoli utenti e/o gruppi di essi diventa più semplice se l'amministratore di sistema può trattenere o distribuire a seconda delle necessità lo spazio disponibile.

Facilità di inserimento di nuovi elementi e migrazione dei dati verso di essi senza discontinuità per l'utente.



Anatomia di LVM

Esempio di organizzazione di LVM su un disco

Hard disk (/dev/sda) 50 GB

Volume fisico
(/dev/sda1) 20 GB

Volume fisico
(/dev/sda2) 30 GB

Gruppo di volumi (/dev/gruppo) 50GB

Volume logico
(/dev/gruppo/vol1) 15 GB

Volume logico
(/dev/gruppo/vol1) 25 GB

Libero
10 GB

Volumi fisici (PV)

Un volume fisico o physical volume (PV) è un hard disk o una partizione di esso, sebbene possa essere anche qualcosa di simile ad esso (un device a blocchi).

Ogni PV è contrassegnato tramite un UUID, che è memorizzato tra i metadati.

Vengono inizializzati mediante il comando `pvcreate` ed è possibile visualizzarne le informazioni tramite `pvdisplay`.

```
[root@john~]# pvcreate /dev/sda1  
[root@john~]# pvcreate /dev/sdb1
```



Gruppi di volumi (VG)

Un gruppo di volumi o volume group (VG) rappresenta un aggregato di risorse.

Ogni gruppo di volumi aggrega tipicamente diversi dischi (o partizioni di essi) in modo da ottenere un'unica sezione di spazio utilizzabile.

Vengono inizializzati mediante il comando `vgcreate` ed è possibile visualizzarne le informazioni tramite `vgdisplay`.

```
[root@john~]# vgcreate gruppo /dev/sda1  
                  /dev/sdb1
```



Volumi logici (LV)

Un volume logico o logical volume (LV) rappresenta una partizione del gruppo di volumi ed è il dispositivo usato per contenere il filesystem.

Vengono inizializzati mediante il comando `lvcreate` ed è possibile visualizzarne le informazioni tramite `lvdisplay`.

```
[root@john~]# lvcreate -L1G -nvol1 gruppo
```





Extent logici (LE) e fisici (PE)

LVM alloca lo spazio di storage in segmenti di uguale grandezza detti extent. La grandezza di tali segmenti può variare tra 8KB e 512MB ed è definita dal gruppo di volumi (VG).

Si parla di extent fisici o physical extent (PE) quando ci si riferisce ai segmenti di dati in cui un volume fisico (PV) è diviso.

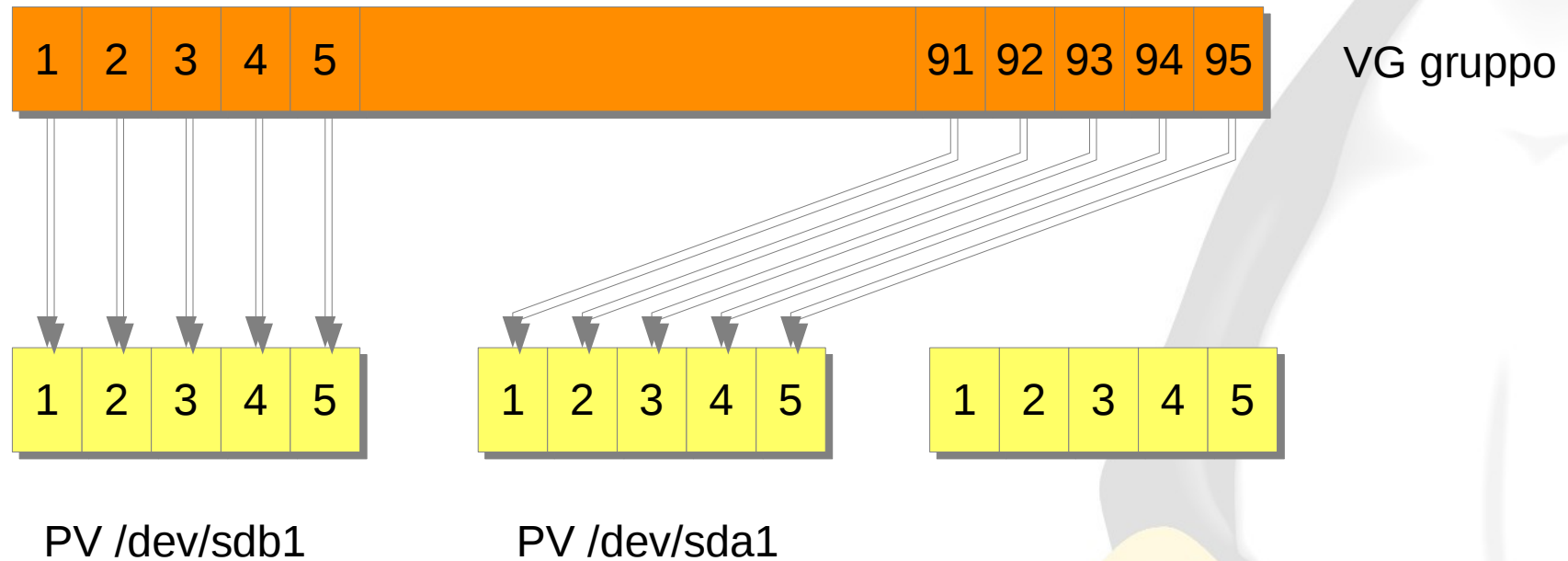
Si parla di extent logici o logical extent (LE) quando ci si riferisce ai segmenti di dati in cui un volume logico (LV) è diviso. Ogni volume logico di un determinato gruppo di volumi ha i LE della stessa grandezza.

A causa di limitazioni del kernel il numero massimo di LE per un VG è limitato a 2^{16} , il che delimita la grandezza massima di un VG (1 EB teorici, 16 TB su sistemi a 32 bit)

Mapping degli extent /1

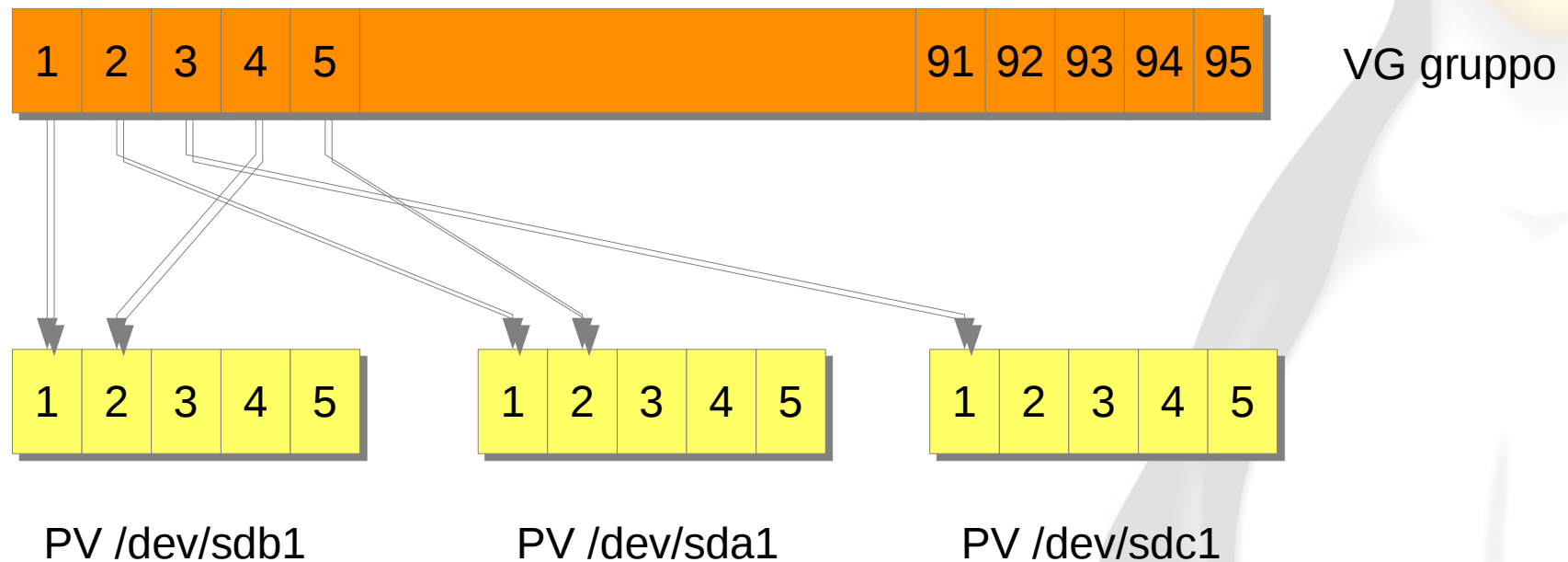
LVM fornisce due strategie per la mappare gli extent logici sugli extent fisici: lineare e striped.

Il mapping lineare assegna un range di LE a un determinato PV



Mapping degli extent /2

Il mapping striped distribuisce i LE su un determinato numero di PV, puntando a migliorare le performance.



Snapshot /1

Creare uno snapshot consiste nel creare un nuovo device a blocchi che rappresenta la copia esatta di un volume logico “congelata” in un preciso momento.

Possiamo comunque continuare a lavorare sul device originale senza preoccuparci della consistenza dello snapshot, questo grazie al meccanismo di copy-on-write.

Il kernel tiene traccia delle modifiche sul device originale e di ciò che è scritto sullo snapshot. Ogni scrittura sul device originale si traduce in

- Lettura dei vecchi dati dal device originale
- Scrittura dei vecchi dati sullo snapshot
- Scrittura di nuovi dati sul device originale



Snapshot /2

È possibile montare gli snapshot sia in modalità sola lettura (read-only) che in modalità lettura e scrittura (read-write), predefinita in LVM2.

È possibile aumentare la grandezza di uno snapshot mentre è montato.

In caso di spazio insufficiente su LV, il kernel abbandona automaticamente lo snapshot.



Manipolazione di VG

Oltre alla creazione tramite `vgcreate`, è possibile rimuovere PV da un VG tramite il comando `vgremove`. Il VG interessato non deve contenere alcun LV.

```
[root@john~]# vgchange -a n gruppo  
[root@john~]# vgremove gruppo
```



L'aggiunta e la rimozione di PV in un VG si ottiene con i comandi `vgextend` e `vgreduce`.

```
[root@john~]# vgextend gruppo /dev/sdc1  
[root@john~]# vgreduce gruppo /dev/sdc1
```



N.B.: prima di rimuovere un PV accertarsi che non sia in uso.

Manipolazione di LV /1

Oltre alla creazione tramite `lvcreate`, è possibile rimuovere PV da un VG tramite il comando `lvremove`. Il LV interessato deve essere chiuso.

```
[root@john~]# umount /dev/gruppo/vol1  
[root@john~]# lvremove /dev/gruppo/vol1
```



Per estendere un LV è necessario usare `lvextend`.

```
[root@john~]# lvextend -L 12G  
/dev/gruppo/vol1
```



Dopo aver esteso il Lv è necessario aumentare la grandezza del file system

Manipolazione di LV /2



La riduzione di un LV è possibile (non per tutti i filesystem) tramite il comando `lvreduce`, ma è bene ricordare che è necessario procedere alla riduzione del filesystem prima del LV.

```
[root@john~]# lvextend -L -1G  
/dev/gruppo/vol1
```



Migrare dati da un PV

Il comando `pvmove` consente di migrare di dati da un PV sia all'interno dello stesso VG

```
[root@john~]# pvmove /dev/sdb
```



Oppure su un nuovo disco inserito all'occorrenza

```
[root@john~]# pvmove /dev/sdb /dev/sdd
```



Fine



Domande?

Linux Day 2007

