RAID

RAID: *redundant array of inexpensive disk*
RAID: *redundant array of indipendent disk*

Il RAID e' una tecnologia per creare un entita' logica unica, composta da dischi singoli a basso costo (di qui il primo acronimo). In effetti e' anche vero che il RAID diviene davvero efficiente se i dischi che vanno a costituire l'unita' fisica sono indipendenti fra loro.
Esistono diversi tipi di configurazioni che hanno prestazioni e fault tollerance differenti.
In effetti un sistema RAID puo' mirare ad incrementare le prestazioni (per esempio scrivendo in parallelo su due dischi si hanno a disposizione due testine e quindi approssimativamente il doppio delle prestazioni[1]) oppure la sicurezza (scrivendo gli stessi dati su due dischi invece che su uno solo, in caso che uno dei due dovesse rompersi, i dati rimarrebbero assolutamente accessibili), oppure entrambi. Ma ora vediamo piu' in dettaglio le configurazioni RAID.

Configurazioni RAID

RAID Lineare

Le partizioni vengono concatenate le une alle altre. Lo spazio disponibile e' pari alla somma della disponibilita' delle partizioni. In pratica questa configurazione viene usata piuttosto di rado. Talvolta usata in RAID software puo' in determinate condizioni essere una scelta migliore di RAID 0 (provare se e' il caso).

RAID 0

Detto anche striping. Le partizioni vengono concatenate e lo spazio viene suddiviso in stripe. I dati e i metadati vengono scritti in chunk scritti su tutte le partizioni, in questo modo migliorando notevolmente I/O grazie all'algoritmo detto round robin.
Le operazioni di I/O vengono spezzate e diventano molte istruzioni eseguite in parallelo (hardware permettendo, ovviamente le partizioni devono trovarsi su diversi controller).
Il problema piu' grande di RAID 0 e' che in caso di fault di uno solo dei dischi, tutto il filesystem viene irrimediabilmente corrotto.

RAID 1

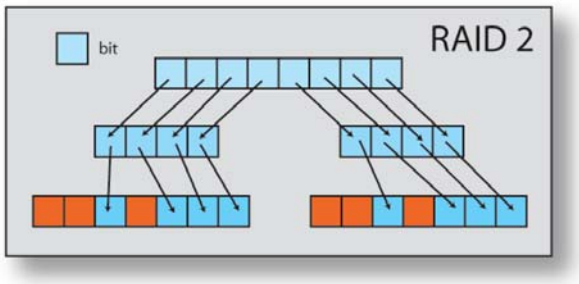
Detto anche mirroring. Per eseguirlo le partizioni devono due di uguale dimensione. I metadata vengono scritti su entrambi i dischi e cosi' i dati veri e propri. In scrittura quindi le prestazioni sono necessariamente non superiori a quelle di un singolo disco. In realta' in lettura si misura un certo miglioramento dovuto al fatto che esse sono eseguite in parallelo.
Rimane il fatto che lo spazio disponibile viene semplicemente dimezzato.

[1] Questo in pratica non e' vero in quanto per una serie di motivi fisici una crescita lineare delle prestazioni rispetto all'aumento dei dispositivi e' abbastanza utopistico

RAID 5

Per potere configurare un RAID 5 ci vogliono N partizioni di uguale geometria e dimensione. Queste vengono combinate in un unico dispositivo logico e lo spazio e' ancora una volta diviso in stripes di blocchi contigui.
Per ogni settore logico del metadevice, si trovano un disco di dati e un altro che contiene la protezione di parita'. In generale i parity blocks sono distribuiti sui dischi attraverso il round robin (in genere).
Per quanto riguarda lo spazio, lo spreco per parita' e' relativamente ridotto. Ricostruire il contenuto di un disco dopo un crash e' fattibile anche se laborioso.
Per quanto riguarda le performance RAID 5 e' molto piu' lento di RAID 1 in lettura, mentre in scrittura le prestazioni sono confrontabili.

RAID 2



Questo tipo di configurazione e' byte-oriented . Ogni byte che compone il flusso dei dati viene scomposto in due segmenti di 4 bit a cui viene aggiunto del codice di hamming cosi che' il primo, secondo e quarto byte sono di parita'. Tutto questo viene scritto su sette hard disk perfettamente sincronizzati. Non e' nemmeno da dire che sia affidabilita' che velocita' teorica sarebbero buone, solo che il costo per avere hardware in grado di sostenere le richieste di Varianti del RAID 2 sono state implementate fino a gestire 39 dischi in parallelo. Decisamente costoso per la maggior parte delle tasche, anche tenendo conto che RAID e' nata come tecnologia relativamente economica, offre comunque prestazioni favolose.

RAID 3

In questo caso si ha solo 1 bit di parita'. Correggere l'errore non e' possibile a meno di conoscere la posizione della corruzione, cosa che comunque avviene in caso di rottura di un disco.

RAID 4

RAID 4 e' basato su stripe, ed in pratica e' un RAID 0 con l'aggiunta del controllo di parita'. Il sistema e' in grado di gestire la rottura su disco e la "perdita" di spazio disponibile e' minima, tuttavia ogni modifica di file comporta la lettura e la riscrittura di tutte le stripe su cui il file si trova, piu' il controllo di parita'. Decisamente lento.

RAID 6

RAID 5 con controllo di parita' bidimesionale (doppio controllo di parita'). Adatto per usi mission critical.

RAID 7

Simile al RAID 4, ma con scrittura asincrona delle stripe. Brevettato dalla Storage Computer Corporation. In realta' il sistema e' molto complesso e piuttosto efficiente, ma e' costoso e proprietario. Evitare.

RAID 0+1

I dati vengono scritti su una delle due partizioni attraverso chunk e stripe e poi sincronizzati.

RAID 1+0

I dati vengono scritti su entrambe le partizioni usando l'interleave del RAID 0.

RAID 53

Due dischi in RAID 0 mirrorati su altri dischi con un altro disco per il controllo della parita'. Questo e' un ottima alternativa al RAID 0+1 in quanto in cambio di un minimo calo di prestazioni e' in grado di gestire la rottura contemporanea di due dischi ai lati opposti del mirror.

RAID 5+1

Mirroring di due volumi RAID 5. Affidabilita' incredibile, prestazioni scarse.

RAID hardware: EIDE

Storicamente RAID e' sempre stato abbinato a SCSI per la possibilita' di questa tecnologia di gestire il TCQ (Tagged Comman Queuing) ovvero di fare eseguire operazioni atomiche in contemporanea a piu' device sullo stesso controller.
La differenza di banda infatti, anche ai primordi, non era abissale (40 MHz dello SCSI contro i 33 dell'IDE), ma un altro forte limite della tecnologia IDE era (e in molti casi e') l'impossibilita' di gestire piu' di sue device per canale, mentre SCSI fin dall'inizio ne supportava 6 (e ora anche di piu').
Poi sono nati controller hardware RAID basati su tecnologia ATA, davvero efficienti, ma in effetti adatti solo a piccoli server, in quanto consentivano esclusivamente RAID 1 e RAID 0.

La rivoluzione di 3Ware

3ware ha di recente presentato de controller ATA RAID con 4-12 canali ATA 66, 100 o 133, indipendenti, con cui e' possibile fare una configurazione RAID hardware, gestita da un chipset. Sulla scheda e' inoltre presente una memory board, che anche se di dimensioni relativamente piccole rispetto a quella dei device SCSI, si rivela piuttosto competitiva in gran parte delle situazioni.
Nonostante la natura della scheda, questa viene vista dall'OS come un device SCSI (e quindi vengono usati i famosi TCQ, che sono controllati da un processore integrato sulla scheda, e permettendo la gestione della periferica attraverso lo SCSI layer (che e' in generale piuttosto performante);

Con questo device e' possibile realizzare RAID 0, 1, 0+1 e 5 ecc, esattamente come su un controller SCSI. Da notare l'ottima qualita' dei driver per il kernel Linux, che consentono il suo utilizzo anche su server fascia media e medio/alta. L'assenza di driver Linux sarebbe stato un grave handicap, in quanto montare un RAID su una workstation Windows e' un po' come montare l'air-bag su una panda, e poi gettarla da un dirupo.

RAID software

Che il Kernel di Linux fosse una struttura estremamente all'avanguardia era risaputo. Che l'architettura di Unix, che identifica ogni dispositivo hardware come un file fosse estremamente efficace era risaputo... Che questo possa portare a dei potentissimi livelli di astrazione anche, e cosi' che il modello di sviluppo a sorgenti liberi fosse estremamente potente. E infatti nessuno si stupisce che lo scheduler del nuovo kernel (per adesso in prova nel 2.5.x e che entrera' nella prossima stable 2.6 o forse gia' 3.0?), cosi' come nessuno si era stupito piu' di tanto quando nel kernel 2.0 avevano fatto le loro timide comparse le patch per inserire nel Kernel il RAID software.
L'idea non era niente affatto male, in quanto consentiva di gestire RAID (anche se non con cosi' tante configurazioni e cosi' tanti dischi come con un controller RAID dedicato). Oggi il supporto a RAID e' integrato nel Kernel, e addirittura le distribuzioni GNU/Linux piu' user-friendly come per esempio Red Hat o SuSE permettono la configurazione del RAID gia' in fase di installazione (purtroppo con molte schede madri non e' possibile mettere /boot su una partizione RAID, ma poco male, tanto in generale su RAID vengono posti /var e /home).
Ovviamente anche in caso di volere gestire il RAID via software, il tutto avviene in modo trasparente all'utente e le performance, specie con dei buoni dischi e una buona macchina, non si discostano molto da un ATA RAID hardware, ma a costo 0.
Rendiamoci conto che una soluzione di questo tipo puo' non essere aliena in un ufficio. Possiamo pensare ad una configurazione hardware del genere.

Canale IDE	**Device**	**Funzione**	**Mountpoint**
Primary Master	/dev/hda	disco di sistema, contiene /, /bin, /usr, /var, /proc, /opt, /mnt e /sbin	/
Primary Slave	/dev/hdb	Questi device verranno messi in RAID 1	
Secondary Slave	/dev/hdd		
Secondary Master	/dev/hdc	masterizzatore	/mnt/cdrw

Bene, dopo avere configurato il RAID 1, i due dischi (si presuppone di uguale dimensione e velocita') saranno visti dal sistema come /dev/md0 (per esempio). Per avere una sicurezza ancora maggiore per i dati, e' possibile utilizzare un filesystem journaled come ext3, che a scapito di un leggerissimo calo di prestazione permette una sicurezza assoluta in caso di crash (non e' possibile che il filesystem si trovi in stato di inconsistenza). Ovviamente una macchina del genere e' adatta per depositare dati, per esempio un fileserver. Infatti il throughput in lettura e' piu' o meno doppio rispetto a quello di un hard disk singolo, e l'eventualita' di perdita dei dati e' piuttosto remota (qualunque sys admin considera la perdita di due dischi in contemporanea un evento piuttosto remoto, anche se da prendere in considerazione in sistemi mission critical). Tutto questo insieme alla capacita' di Linux di gestire un numero elevato a piacere di

connessioni (a differenza di altri sistemi operativi che le limitano in base alla licenza), puo' creare un repository davvero efficiente, anche perche' al giorno d'oggi un hard disk IBM 7200rpm 120 GB costa circa 130 euro (se si e' disposti a girare un po' di negozi, lo si trova anche a meno).

RAID Hardware: SCSI

La domanda che a questo punto dobbiamo porci e': abbiamo ancora bisogno di controller SCSI? La risposta puo' essere duplice. Da un lato EIDE si avvicina sempre piu' come stabilita' e velocita' a SCSI, e la sua ampia diffusione la rendono decisamente economica, dall'altro una grande azienda puo' trascurare il costo maggiore di periferiche SCSI, in nome di maggiore affidabilita'.
I tempi in cui SCSI volava spanne sopra EIDE (che poi coincidevano con i tempi in cui nei Mac di ogni fascia venivano montati dischi SCSI) sono ormai trascorsi (al momento attuale anche su un dual-G4 si trova un Ultra-ATA).
Inoltre ci sono tecnologie come serial-ATA che spingono il mercato. D'altra parte molte tecnologie tradizionalmente associate a SCSI (i vecchi scanner per esempio) ormai usano USB o USB2 (che con i suoi 480Mb/s raggiunge una velocita' di tutto rispetto).
Non so quanto spazio possa rimanere per SCSI in ambito consumer, ma d'altra parte computer quali i server mainframe IBM non abbandoneranno per lungo tempo questa tecnologia (anzi sono arrivati a connettere dispositivi SCSI tramite fibra ottica).
Inoltre il RAID SCSI permette di avere prestazioni eccezionali (512 MB di cache per I/O e' davvero impressionante) e permettono anche di gestire una trentina di hard disk (un computer di questo tipo era rimasto per un po' all'Universita' di Pisa, e faceva uso di una variante del RAID 2)..
Tuttavia ormai sul mercato girano server da uno o piu' Terabyte di capacita' (si ricorda che possedendo un TB di mp3 illegali si puo' essere fucilati dalla SIAE sulla pubblica piazza, a prescindere che -- per fortuna-- in Italia non ci sia la pena di morte), con prestazioni di 21 MB/s in scrittura e 75MB/s in lettura con files di 4 GB (ma raggiungendo senza colpo ferire i 300 MB/s con files da "1 solo" GB), basati su tecnologia EIDE, del tutto accessibili a prezzi (relativamente) contenuti.

Alcuni consigli conclusivi

RAID non e' volto a sostituire il backup, serve a rendere un data crash meno probabile (oppure addirittura ad aumentare semplicemente le prestazioni). Backuppare il sistema (cosa perfettamente legale, nonostante l'EUCD abbia introdotto il balzello per la SIAE su ogni dispositivo che potenzialmente potrebbe essere usato a memorizzare dati che violano il diritto d'autore) e' importante in ogni caso (e lo dico dopo che 5 giorni fa uno dei miei hard disk ha semplicemente deciso che non funzionare fosse un suo diritto, purtroppo proprio mentre stavo facendo le iso per il backup :-( ).
Inoltre mettere in mirroring due device sullo stesso canale/controller ATA/SCSI e' sconsigliato, in quanto il fault di uno, potrebbe bloccare l'intero canale, rendendo inaccessibile anche il device sano.
Sembra stupido da dire, ma spegnere sempre il computer in modo pulito e' assolutamente consigliabile (specie con RAID 0).

Un altro consiglio banale e' quello di non mettere nello stesso array di dischi due partizioni sullo stesso hard disk. Infatti in questo caso non si incrementano ne' le performance, ne' il fault tollerance del sistema (anzi probabilmente lo si diminuisce).
Se cercate un sistema davvero performante scegliete senza dubbio hardware RAID. Anche se in alcuni casi software RAID ha dimostrato performance superiori, questo e' vero solo in determinate situazioni.
Infine non trattero' per nulla quale possa essere la migliore configurazione riguardo a dimensioni di chunk e stripes a seconda del tipo di server, in quanto il discorso e' piuttosto lungo, ed in ogni caso dovrei appoggiarmi a benchmark di altri, in quanto non ho modo di svolgere i test io stesso. Pertanto consiglio di accedere tali benchmark direttamente da internet.

WEBOGRAPHY

AC&NC: schemi dei vari livelli RAID, caratteristiche. Non ho incluso gli schemi (anche se belli) in quanto sotto copyright)
Apple Raid: praticamente lo stato dell'arte
1U-RAID 5: un sito dedicato a RAID
Anubi: un sito di informatica, non strettamente legato a RAID, ma potrete scaricare versioni aggiornate di questo documento
Vecchi documenti: al momento sono obsoleti ma utili a capire cosa era RAID anni fa [tedesco]
Linux Journal: un articolo particolarmente tecnico, ma anche estremamente istruttivo (anche se ormai obsoleto ^__^)
IBM: alcune notizie sul RAID di Linux (particolarmente interessante in quanto in breve BigBlue distribuira' i suoi server esclusivamente con GNU/Linux)