**I/O**

**Differenza tra SRAM e DRAM?**

Le DRAM utilizzo un solo transistor per bit mentre le SRAM utilizzano da 6-8 transistor a bit (per evitare che l’informazione possa essere disturbata quando viene letta) quindi le DRAM sono molto più dense rispetto ad una SRAM inoltre le DRAM rispetto alle SRAM hanno bisogno del refresh, per effetuare il refresh di una cella occorre semplicemente leggerne il contenuto e riscriverlo.

**Che cos’è il Direct Memory Access?**

**DMA** un meccanismo che fornisce al controllore di un dispositivo la possibilità di trasferire direttamente alla/dalla memoria senza coinvolgere il processore.

Sia per il trasferimento governato dal polling e sia per quello governato dagli interrupt, l’onere di gestire ed eseguire il trasferimento dei dati ricade sul processore.

Per i dispositivi a banda larga come i dischi rigidi il trasferimento riguarda soprattutto blocchi di dati relativamente grandi, quindi è stato inventato un meccanismo per scaricare il processore e far sì che il controllore del dispositivo trasferisca i dati direttamente in memoria senza coinvolgere il processore.

Questo meccanismo è chiamato accesso diretto alla memoria DMA.

Il DMA è implementato attraverso un controllore specializzato che trasferisce i dati tra un dispositivo di I/O e la memoria indipendentemente dalla memoria.

Il controllore di DMA diventa master della connessione e gestisce la lettura e la scrittura dei dati tra il dispositivo e la memoria.

Ci sono tre fasi in trasferimento DMA

1. Il processore inizializza il DMA fornendo l’identità del dispositivo. L’operazione che esso deve eseguire è quella di trasferire l’indirizzo sorgente o destinazione dei dati e il numero di byte.
2. Il DMA inizia l’operazione sul dispositivo e governa la connessione. Quando il dato p disponibile lo trasferisce. Il DMA fornisce l’indirizzo di memoria per la singola lettura o scrittura. Se la richiesta implica più di un trasferimento, l’unità di controllo genera l’indirizzo successivo e inizia il relativo trasferimento del dato. Utilizzando questo meccanismo l’unità DMA può completare un intero trasferimento, che può essere lungo miglia di byte, senza far intervenire il processore.
3. Una volta terminato il trasferimento DMA, il controllore può interrompere il processore, il quale può verificare se l’intera operazione sia stata completata con successo.

Nello stesso calcolatore ci possono essere dispositivi DMA multipli.

**Dependability? Come si misura l’affidabilità?**

È una misura della continuità con cui viene fornito il servizio a partire da un certo istante, ossia è il tempo che intercorre prima che si verifichi il primo malfunzionamento.

Erogazione del servizio, interruzione del servizio.

**RAID**

**Che cos’è un disco RAID?**

Un’unità disco strutturata come un insieme di dischi piccoli e poco costosi, in modo da migliorare prestazioni e affidabilità.

I RAID sono nati perché prima un disco di grandi dimensioni era assai costoso quindi si è pensato di utilizzare più dischi di piccole dimensioni.

Avere più dischi invece di un solo comportava uno svantaggio; se ho n dischi la probabilità di guasto va moltiplicata per n il che significa che l’affidabilità di tutto il gruppo di dischi si abbassa.

Per sopperire a questo problema l’idea era quella di utilizzare dell’opportune distribuzioni di dati e dei codici per la correzione degli errori in maniera tale da poter distribuire i dati tra i vari dischi assicurando però una affidabilità uguale se non superiore a quella di un unico disco di grandi dimensioni.

**Perché scelgo RAID?**

Con l’utilizzo dei dischi RAID si ha un miglioramento delle prestazioni perché ci sono unità che possono lavorare in parallelo e quindi posso andare più veloce, inoltre migliora sicuramente l’avaiability (tempo per cui la risorsa è utilizzabile).

Chiaramente per avere un avaiability alta quindi non bisogna mai spegnere il sistema è necessario l’hot swapping.

Scelgo i dischi RAID perché garantiscono una maggiore affidabilità ottenibile grazie alla ridondanza dei dischi.

Con i dischi RAID è possibile costruire un sistema fault tollerance cioè un sistema che nonostante i guasti continui a funzionare.

Il fault tollerance è particolarmente interessante se i dischi guasti sono hot swapped.

**Chi utilizza i dischi RAID?**

I dischi RAID vengono utilizzati soprattutto nei server perché migliorano le prestazioni (aumentano la banda passante in modo da fornire migliori prestazioni) ma la motivazione fondamentale per cui utilizzo i RAID è che permette di fare cose che nessun altro fa.

Se un disco si guasta non si spegne la macchina semplicemente si sostituisce il disco con un altro, il sistema nella sua interezza continua a funzionare anche se un po’ più lentamente; dopodiché si lancia un programma che ricostruisce quello che manca sul disco.

Nel senso che la ridondanza è tale per cui senza un disco si continua ad avere accesso a tutti i dati e senza un disco è possibile ricostruire tutta la situazione senza spegnere la macchina

**Diverse strutture dei dischi RAID**

Esistono vari livelli tra i dischi RAID quello che cambia tra i vari livelli sono l’affidabilità la velocità e soprattutto il numero di dischi.

RAID 0 (Nessuna ridondanza)

I dati vengono suddivisi su più dischi (striping) così facendo si forza l’accesso a dischi diversi.

Lo striping su un insieme di dischi fa si che il disco RAID appare al software come un unico grande disco e in più migliora le prestazioni per accesi a dati di grandi dimensioni poiché molti dischi possono lavorare contemporaneamente.

I RAID 0 non contengono ridondanze

RAID 1 (Mirroring)

Schema tradizionale utilizzato per compensare i guasti sui dischi viene chiamato mirroring, utilizza il doppio dei dischi della struttura RAID 0, ogni volta che un dato viene scritto su un disco, lo stesso dato viene scritto anche sul disco ridondante, cosicché esistono sempre due copie della stessa informazione e se un disco subisce un guasto si accede al disco mirror.

RAID 2 (Riconoscimento degli errori e codice di correzione degli errori)

Suddividere a livello di byte o di blocchi i dati tra i dischi utilizzando poi a fette un codice per il controllo dell’errore in maniera tale che questo codice per il controllo possa servire a rigenerare i dati che si perdono.

L’idea alla base dei vari livelli è sempre la stessa utilizzare un codice per la correzione dell’errore che permette di far sopravvivere i dati che in orizzontale sono stati spalmati o a livello di byte o a livello di blocchi fra i vari dischi, in questo cambiano i vari livelli.

È importante non avere un disco con tutti codici di correzione dell’errore perché qualsiasi scrittura mi coinvolge sempre lo stesso disco i codici per la correzione dell’errore si mettono in diagonale in maniera tale da distribuire gli accessi.