

### Consideraciones en los sistemas

Cuando el almacenamiento de disco resulta un cuello de botella, se tiende a agregar más discos al sistema. Hacerlo no siempre resuelve el problema porque el cuello de botella puede deberse a una carga pesada de solicitudes sobre un número relativamente pequeño de discos. Si se detecta esta situación, la planificación de disco puede servir para mejorar la eficiencia y eliminar el cuello de botella.

**Memoria Caché de Disco.-** Es un área de almacenamiento primario en la que se conservan los registros de acceso frecuente para ayudar a evitar la necesidad de búsquedas largas de disco.

Cuando se realiza una operación de escritura, podría pensarse que el registro se graba de inmediato en el disco. Así sucede en algunos sistemas, pero en otros con memoria caché de disco la escritura sólo hace que el registro se almacene en un buffer en almacenamiento primario; el registro permanece ahí hasta que el sistema se queda sin espacio de buffer para escrituras subsecuentes, y en ese momento se graba el registro en el disco.

Si fuera necesario leer un registro escrito recientemente, se puede obtener del buffer de memoria caché de disco en almacenamiento primario mucho más rápido que si tuviera que leer del disco.

La clave para aprovechar la memoria caché de disco es mantener los registros de acceso frecuente en el buffer de memoria caché de disco de almacenamiento primario. Desde luego esta técnica sólo funciona bien cuando es posible identificar los registros de acceso frecuente. Utilizando la heurística de localidad, un registro al que se hizo referencia en el pasado reciente talvez será requerido en el futuro cercano.

### Otras técnicas para mejorar el desempeño

La optimización del desempeño de los dispositivos rotacionales de almacenamiento se ha tratado de lograr empleando métodos de hardware, de sistemas operativos y de sistemas de aplicación. Por ello podríamos decir que otras técnicas para mejorar el desempeño son:

- La reorganización del disco para reducir la fragmentación;
- La partición del disco en la que se confinan los archivos a zonas pequeñas a fin de reducir la fragmentación;
- Colocar varias copias de datos estables de referencia frecuente en muchas partes del disco para reducir las distancias de búsqueda;
- La duplicación de datos estables de referencia frecuente en unidades de disco de acceso individual para lograr una mayor concurrencia;
- La agrupación de registros en bloques para reducir el número de búsquedas;
- El mantenimiento de datos de acceso frecuente en posiciones de acceso más rápido dentro de la jerarquía de almacenamiento;
- La colocación de datos en las pistas de la zona media en dispositivos que utilizan una planificación tipo SCAN;
- La lectura de una pista completa a la vez para aprovechar la localidad espacial
- La compresión de datos para reducir el espacio requerido y por consiguiente, los tiempos de acceso

En general se ha evitado el aumento en las velocidades de rotación de los discos por causa de limitaciones físicas.

**Discos de Ram.-** Es un dispositivo de disco simulado en memoria convencional de acceso aleatorio. Elimina por completo los retrasos que se dan en los discos convencionales a causa de los movimientos mecánicos inherentes a las búsquedas y a la rotación del disco (los discos en RAM no implican movimientos mecánicos). Son útiles sobre todo en las aplicaciones de alto rendimiento.

Están separados de la memoria principal, de modo que no utilizan espacio requerido por el sistema operativo o las aplicaciones. Los tiempos de referencia a datos individuales son uniformes, sin la amplia variabilidad de los discos convencionales.

Los discos en RAM son caros. Son volátiles, es decir pierden su contenido cuando se apaga la computadora o se interrumpe el suministro de energía, podríamos solucionar esto con el uso de baterías de respaldo (UPS), pero si el corte de energía se prolonga estas se pueden agotar por lo que siempre es más recomendable hacer respaldos frecuentes en discos convencionales.

**Discos de Opticos.-** Pueden contener enormes cantidades de datos bajo un régimen de grabación-única-lectura-múltiple o WORM write-once-read-many (los primeros discos láser). Los discos láser regrabables podrían desplazar los discos magnéticos de cabeza móvil.

Resultan apropiados para aplicaciones de archivo, pero no son muy útiles para aplicaciones que requieren actualización regular. Las capacidades enormes de estos discos (varios gigabytes, quizá 100 veces mayor que los discos duros empleados normalmente en las PC's) han hecho que resulten realmente efectivos en algunas aplicaciones

## RELOJES

También llamados cronómetros son esenciales para la operación de cualquier sistema de tiempo compartido. El software de reloj toma la forma de un manejador de dispositivo, aunque un reloj toma por la forma de un manejador de dispositivo, aunque un reloj no es un dispositivo de bloque, como un disco, ni un dispositivo de carácter, como una terminal.

### Cronómetro de intervalos o reloj de interrupciones

Los relojes no tienen direcciones por medio de bloques, tampoco generan o aceptan flujos de caracteres, lo único que hacen es provocar interrupciones a intervalos bien definidos.

Un proceso que tiene asignado el CPU está en ejecución. Si el proceso pertenece al sistema operativo, se dice que el sistema operativo está en ejecución y puede tomar decisiones que afectan la operación del sistema. Para evitar que los usuarios monopolicen el sistema (accidental o deliberadamente), el sistema operativo tiene mecanismos para arrebatar el CPU al usuario.



El sistema operativo mantiene un *reloj de interrupciones o cronómetro de intervalos* para generar interrupciones en algún momento futuro específico (o después de cierto tiempo). El reloj de interrupciones ayuda a garantizar tiempos de respuesta aceptables para los usuarios interactivos, evita que el sistema quede bloqueado en un ciclo infinito de algún usuario y permite que los procesos respondan a eventos dependientes de tiempo. Los procesos que deben ejecutarse periódicamente dependen del reloj de interrupciones.

Por lo general las labores de un reloj son:

1. Mantener la hora del día
2. Evitar que los procesos se ejecuten más tiempo del permitido.
3. Mantener un registro del uso del CPU
4. Controlar la llamada al sistema ALARM por parte de los procesos del usuario.
5. Proporcionar cronómetros guardianes de partes del propio sistema.
6. Realizar resúmenes, monitoreo y recolección de estadísticas

### TERMINALES

Cada computadora tiene una o más terminales que se utilizan para comunicarse con ella. Las terminales tienen un gran número de formas distintas. El manejador de terminal se encarga de ocultar todas estas diferencias, de forma que la parte independiente del dispositivo en el sistema operativo y los programas de usuario no tienen que volverse a escribir para cada tipo de terminal.

Los editores de pantalla y muchos otros programas sofisticados deben poder actualizar la pantalla de maneras complejas que le simple recorrido del texto en la parte inferior de la misma. Para lograr esto, muchos manejadores de terminales soportan varias secuencias de escape. A continuación se listan algunas:

1. Mover el cursor hacia arriba, abajo, a la izquierda o a la derecha una posición.
2. Mover el cursor a  $x, y$
3. Insertar un carácter o una línea  $n$  en el cursor
4. Eliminar un carácter o una línea del cursor
5. Recorrer la pantalla hacia arriba o hacia abajo  $n$  líneas
6. Limpiar la pantalla desde el cursor hasta el final de la línea o hasta el final de la pantalla.
7. Trabajar en modo video inverso, subrayado, parpadeo o normal
8. Crear, destruir, mover o controlar ventanas.

Cuando el manejador ve el carácter que inicia la secuencia de escape, activa una bandera y espera a que llegue el resto de la secuencia. Cuando toda la secuencia ha llegado, el manejador debe llevarla a cabo en software. La inserción y eliminación de texto requiere el movimiento de bloques de caracteres en el video RAM. El hardware no ayuda más que en recorrer descendente o ascendentemente texto y en exhibir el cursor.

### DESEMPEÑO, COPROCESADORES, RISC Y FLUJO DE DATOS

Tres objetivos comunes de la evaluación del rendimiento:

- 1) **Evaluación para la Selección.**- El evaluador del desempeño debe decidir si conviene adquirir un sistema de cómputo de un proveedor específico.

- 2) **Proyección del Rendimiento.**- El objetivo del evaluador en este caso es estimar el desempeño de un sistema inexistente. Puede tratarse de un sistema de cómputo totalmente nuevo ó de un nuevo componente de hardware o software.
- 3) **Supervisión del Desempeño.**- el evaluador acumula datos de un sistema ó componente ya existente para asegurarse de que cumple con sus objetivos de desempeño, para ayudar a estimar el impacto de las modificaciones y para ofrecer a los administradores la información que necesitan para tomar decisiones estratégicas tales como modificar o no un sistema existente de prioridades de trabajo.

**Evaluación y Predicción del Desempeño.**- Se necesitan desde los primeros momentos de la concepción de un nuevo sistema, en la operación cotidiana del sistema después de la instalación y cuando se estudia la modificación o posible sustitución de un mejor sistema.

**Desempeño.**- Es la eficiencia con la que un sistema cumple sus objetivos

#### Medidas de desempeño

**Tiempo de Retorno.**- En un sistema de procesamiento por lotes se define como el tiempo transcurrido desde la entrega de un trabajo hasta la devolución al usuario.

**Tiempos de Respuesta.**- Es el tiempo de retorno en un sistema interactivo y a menudo se le define como el tiempo transcurrido desde que el usuario presiona el "enter" ó "botón del mouse" hasta que el sistema comienza a imprimir o exhibir una respuesta.

**Tiempo de Reacción del Sistema.**- Es el tiempo transcurrido desde que el usuario presiona "enter" ó el "botón del mouse" hasta que se otorga la primera tajada de tiempo de servicio a solicitud del usuario.

**La Varianza en los Tiempos de Respuesta.**- Es una medida de predecibilidad.

**Producción.**- Es la medida del rendimiento de trabajo por unidad de tiempo

**Carga de Trabajo.**- Es la medida de la cantidad de trabajo que se ha introducido al sistema, cantidad que el sistema debe procesar en condiciones normales para que su funcionamiento se considere aceptable.

**Capacidad.**- Es la medida de producción máxima que puede tener un sistema.

**Utilización.**- Es la fracción de tiempo en que está en uso un recurso (es una medida engañosa porque al parecer lo mejor es tener un porcentaje de utilización alto y ésto podría ser señal de un aprovechamiento ineficiente).

#### Técnicas para Evaluar el Desempeño

**Tiempos.**- Son útiles para realizar comparaciones rápidas entre equipos.

**Mezcla de Instrucciones.**- Emplean un promedio ponderado de diversos tiempos de



instrucciones más adecuadas para una aplicación específica.

**Programa Núcleo.-** Es un programa representativo que podría ejecutarse en una instalación. Se cronometra para una máquina dada empleando las estimaciones de tiempo de instrucciones del fabricante y así se pueden hacer comparaciones entre máquinas distintas de acuerdo con la velocidad de ejecución esperada del programa núcleo.

**Modelos Analíticos.-** Son representaciones matemáticas de sistemas de cómputo o de sus componentes. Existe un volumen considerable de resultados matemáticos que puede aplicar el evaluador para ayudar a estimar el desempeño de un sistema de cómputo dado (Ejemplo: Teoría de colas y modelo de Markov).

**Bancos de Prueba.-** Es un programa real que el evaluador ejecuta en un sistema de cómputo en evaluación. El evaluador conoce las características del rendimiento del banco de prueba en un equipo ya existente de tal modo que cuando se ejecuta en un equipo nuevo el evaluador puede sacar conclusiones significativas.

**Programas Sintéticos.-** Son programas reales diseñados a la medida para ejercitar funciones específicas de un sistema de cómputo. Son útiles sobre todo cuando no existen bancos de prueba que realicen dichas funciones.

**Simulación.-** Es una técnica con la cual el evaluador desarrolla un modelo computarizado del sistema de evaluación. Luego el modelo se ejecuta en un sistema de cómputo con lo que se refleja el comportamiento del sistema en evaluación.

- Hay simuladores que se manejan por eventos que se producen en el simulador de acuerdo con distribuciones de probabilidades.
- Los simuladores manejados por libreto son controlados por datos derivados empíricamente y manipulados con cuidado de manera que reflejen el comportamiento esperado del sistema simulado.

**Supervisión del Desempeño.-** Es la obtención y el análisis de información acerca del rendimiento de sistemas ya existentes.

#### **Cuellos de Botella y Saturación**

**Cuello de Botella.-** Un recurso se convierte en *cuello de botella* y limita el desempeño total del sistema cuando no es capaz de manejar el trabajo que se le envía.

**Saturación.-** Los recursos que operan cerca de su capacidad máxima tienden a saturarse, es decir, los procesos que compiten por la atención del recurso comienzan a interferir unos con otros (Ejemplo: Hiperpaginación)

**Ciclo de Retroalimentación.-** Es un caso en el que cierta información acerca del estado

actual del sistema puede afectar las solicitudes que llegan. Si la retroalimentación indica que dichas solicitudes pueden tener problemas para ser atendidas, quizás se les pueda enviar por otro conducto.

**Retroalimentación negativa.-** La tasa de llegada de retroalimentación negativa contribuye a la estabilidad en sistemas manejados por colas (La tasa de llegada de solicitudes nuevas puede disminuir como resultado de la información retroalimentada).

**Retroalimentación positiva.-** En estos sistemas, la información retroalimentada origina un aumento en algún parámetro, puede provocar inestabilidad en sistemas manejados por colas.

**Coprocadores.-** Son procesadores de aplicación especial que se agregan a los sistemas de cómputo para realizar operaciones que no están incluidas en el procesador o procesadores originales.

**RISC .-** Las arquitecturas de un RISC casi siempre tiene conjuntos de instrucciones en lenguaje de máquina poco abundantes (Computación con un conjunto reducido de instrucciones) relativamente sencillas de carga y almacenamiento para transferir datos entre la memoria y los registros con ductos profundos y memoria caché. Aunque los programas RISC suelen ser mas tardos que sus equivalentes en CISC, casi siempre se ejecutan en forma rápida.

#### **COMPUTACIÓN DISTRIBUIDA: LA PERSPECTIVA DE LA COMPUTACIÓN EN PARALELO.**

**Canalización.-** Es una técnica para mejorar el rendimiento permitiendo que varias instrucciones en lenguaje de maquina estén en diferentes etapas de ejecución al mismo tiempo. Cada etapa de la canalización de instrucciones realiza una etapa diferente sobre la instrucción y después la instrucción avanza a la siguiente etapa. La canalización hace posible que muchas instrucciones, cada una en diferente fase de ejecución, progresen al mismo tiempo.

**Multiprocesadores.-** Uno de los atractivos de un sistema con múltiples procesadores o multiprocesamiento es que si falla un procesador por lo regular se puede seguir trabajando con los procesadores restantes. El sistema operativo debe darse cuenta de que un procesador específico ha fallado y ya no está disponible para ser asignado.

Hay dos técnicas comunes para detectar el paralelismo:



**1. Distribución de ciclos:** Considere la siguiente instrucción que suma los elementos de los arreglos "b" y "c" y coloca las sumas en el arreglo "a". Este ciclo hace que un procesador secuencial realice las 4 interacciones del ciclo una tras otra.

```

for i = 1 to 4 do
  a(i) = b(i) + c(i)
cobegin
  a(1) = b(1) + c(1)
  a(2) = b(2) + c(2)
  a(3) = b(3) + c(3)
  a(4) = b(4) + c(4)
coend

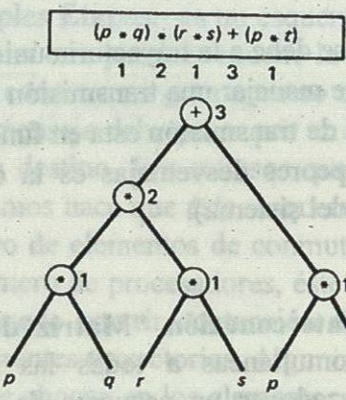
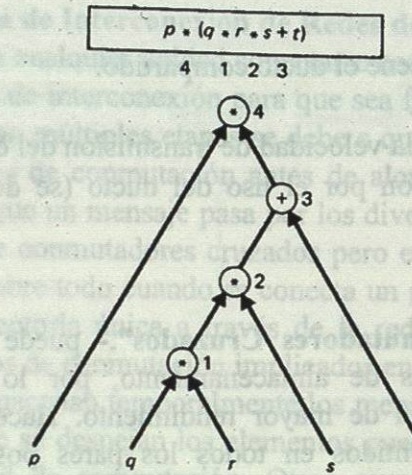
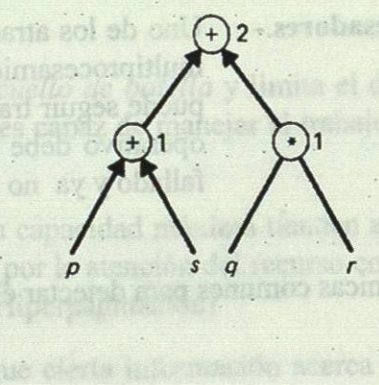
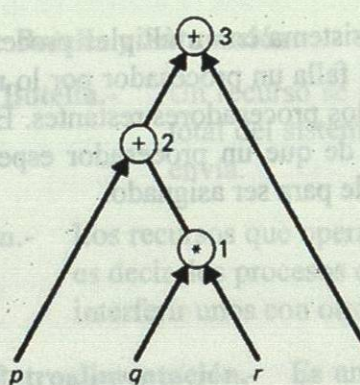
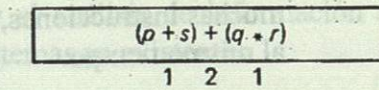
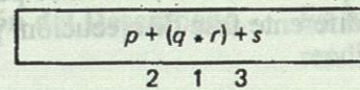
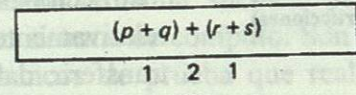
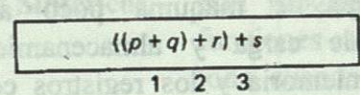
```

Todas las interacciones se pueden realizar independientemente una de otra separándose en unidades ejecutables de manera concurrente (al mismo tiempo).

**2. Reducción de Altura de Arboles.-** Aplica las propiedades de conmutatividad, asociatividad y propiedad distributiva de las operaciones aritméticas para reacomodar las expresiones algebraicas y hacerlas más adecuadas para la ejecución concurrente.

$$((p + q) + r) + s$$

$$(p + q) + (r + s)$$



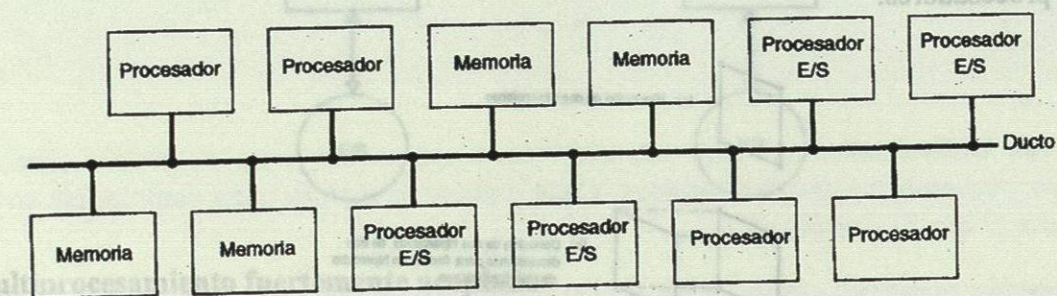
**Regla de Nunca Esperar.-** Establece que es mejor darle al procesador una tarea que puede ser utilizada o no posteriormente que dejar al procesador sin trabajo.

**ESQUEMAS DE INTERCONEXION DE LOS PROCESADORES**

Los problemas fundamentales en el diseño de sistemas de multiprocesamiento es determinar la forma de conectar los múltiples procesadores y los procesadores de E/S con las unidades de almacenamiento.

**Esquema de interconexión "Ducto Compartido".-** utiliza una sola trayectoria de comunicación con todos los procesadores, las unidades de almacenamiento y los procesadores de E/S. Esencialmente es una unidad pasiva.

Ejemplo: Esquema de red de área local: Ethernet.



**Ventajas:**

- Se le pueden fácilmente adicionar nuevas unidades, basta conectarlas directamente al ducto para que ocurra la comunicación, cada unidad debe saber cuales otras unidades están conectadas al ducto, esto se maneja por software.