

System Performance Evaluation Corp. (SPEC)

Antes del SPEC

Existía como parámetro: Velocidad de reloj, o cantidad de instrucciones procesadas por unidad de tiempo MIPS, se convierte en un problema cuando se compara los RISC. Usan más instrucciones que los CISC para un mismo trabajo.

Benchmarks sintéticos: Whetstone y Dhrystone.

Eran una mezcla resultante del estudio de aplicaciones, que comienzan a quedar excluidos con el uso de compiladores optimizados, además podrían llegar a arrojar valores convenientes en aquellos sistemas con características relacionados con el benchmark por ejemplo sobrecargado en el uso de funciones.

SPEC (resultado de la unión de DEC, IBM, SUN, HP; Intel) compañía sin fines de lucro.

Están divididos en prueba de cálculos enteros y de punto flotante. SPEC89/SPEC92 más reciente amplía la cantidad de programas de cada rama de 10 a 20. Ambos grupos miden la performance de cada programa y combinan los valores en estadísticas resumidas, tomando tiempos de ejecución y calcular el SPECratio que divide por el valor de referencia por el tiempo de ejecución. El valor de referencia se toma respecto de una VAX-11/780.

SPEC89 definía tres relaciones métricas

SPECint89 = promedio geométrico de los SPECratio

SPECfp89 = promedio geométrico de los SPECratio

SPECmark89 = promedio geométrico de SPECint89 y SPECfp89

SPEC92 no toma los SPECmark y se redefinen para medir la eficacia con la que el sistema maneja las operaciones multitasking y se calculan ejecutando varias copias de un benchmark de manera simultáneas:

SPECrate_fp92

SPECrate_int92

La fórmula SPECrate toma la relación de tiempo de referencia contra tiempo medido y obtiene una escala que multiplica por un valor constante y por el número de veces que se ejecuta el benchmark, no comparable con los SPECfp o SPECint, permiten comparar como una arquitectura versus otra se degrada debido a las operaciones multitasking.

SPEC92 contiene 14 programas para punto flotante y 6 para entero. Estos se publican en el boletín técnico SPEC y se toman del web de Univ.de Tennessee.

Cómo se interpreta SPEC.

Están orientados fuertemente a medir la performance del procesador (la habilidad del procesador, sistema de memoria y compilador).

Los mainframe brindan un valor modesto debido a que están orientados a realizar operaciones de E/S, si lo reemplazamos por una workstation con valores superiores produciríamos un desastre.

Se debe saber qué se mide.

La performance se limita al eslabón de la cadena más lento. Un motivo común de demoras por ejemplo, son las operaciones de E/S, éstas se activan cuando el sistema operativo se queda sin RAM y se ve obligado a intercambiar datos con el exterior. Para muchos sistemas la duplicación de RAM sería más importante que duplicar la velocidad del procesador.

Una solución más real podría ser reunir sus propios resúmenes estadísticos eligiendo los miembros del conjunto SPEC que tenga mayor relación con dichos códigos e ignorando los restantes.

Algunas arquitecturas son especialmente buenas o malas para el manejo de un nivel de precisión o del otro.

La vectorización también puede dar resultados marcadamente dispares; una arquitectura de vector o una superescalar con un buen compilador ejecutará rápidamente códigos vectorizables. Si se puede vectorizar el programa que se tiene lo crítico recae sobre la arquitectura y el compilador, por el contrario si no se puede los valores de SPECfp serán confusos.

El resumen propio estadístico se lo denomina MySPECfp la manera es tomar aquellos que sean similares a las aplicaciones propias e ignorar el resto. Otra forma es asignar a cada aplicación un valor basados en el grado de relevancia y luego calcular el valor MySPECfp y se usa este para efectuar la comparación. También se puede calcular la relación de MySPECfp contra SPECfp.

Los SPEC no representan bien las *operaciones con punteros*, así si mi aplicación utiliza intensivamente estas operaciones los valores *SPEC no serán representativos*.

Tampoco utilizan la *división fuerte de enteros*, que si es utilizada para el manejo de imagen comprimidas con formato *MPEG*, en este caso *utilizar otros test*.

SPEC está introduciendo una nueva clasificación, que se denomina **SPECbase**, que incorpora un conjunto de restricciones en los identificadores (flags) que se pueden especificar durante la compilación, con otra denominación SPECbase_fp92.,etc.

Necesita que usen los mismos flags para todos los benchmarks y que las opciones sean seguras, se deberán especificar los SPECbase o bien los resultados de SPECbase y los resultados de precisión completos.

AIM

Controla rigurosamente los benchmarks de su Suite VI y Suite VII. Estos generan informes para cada combinación de tareas:

- pico de performance
- performance sostenida
- índice de sincronización de trabajos

El *pico de performance* mide simplemente el número máximo de trabajos por minutos que la computadora puede manejar.

La *performance sostenida* muestra el máximo nivel de performance que un sistema particular puede alcanzar antes de empezar a sufrir caídas abruptas de performance (por ejemplo una causa forzada antes de que el sistema pueda atender un proceso).

El *índice de sincronización* de trabajos muestra diferencias potenciales en el tiempo que tardan en completarse los trabajos concurrentes (se busca lo más cercano a 100).