

SESIÓN DE LABORATORIO 10

El benchmark SPEC CPU2017

Objetivos

- Entender qué aspectos específicos del sistema informático evalúa el benchmark CPU2017 desarrollado por la corporación SPEC.
- Conocer el tipo y la naturaleza de los programas que integran el benchmark SPEC CPU2017 y los criterios que se han aplicado para seleccionarlos.
- Comprender la forma en que se calculan los índices de rendimiento publicados en la web del benchmark SPEC CPU2017 y las implicaciones que ello conlleva.
- Analizar con detalle los resultados de una evaluación de rendimiento de un servidor llevada a cabo con el benchmark SPEC CPU2017.

Desarrollo

El primer objetivo de esta sesión de laboratorio es entender qué es, para qué sirve y qué criterios se han seguido para diseñar el benchmark SPEC CPU2017. En particular, interesa conocer qué tipo de programas lo componen y cómo se organizan dentro de la aplicación.

Adicionalmente, se utilizarán los resultados obtenidos con este programa de prueba, los cuales son públicos y están disponibles en la web de la corporación SPEC, en un servidor real de la serie HP ProLiant de Hewlett Packard para comprender la manera en que se calculan los índices de rendimiento y poder interpretarlos adecuadamente en estudios de comparación de prestaciones.

El benchmark SPEC CPU2017

Para responder correctamente a las siguientes cuestiones es necesario leer detenidamente la información sobre SPEC CPU2017 publicada en la siguiente dirección URL de SPEC:

- <https://www.spec.org/benchmarks.html#cpu>

En particular, resulta de especial interés el contenido de la página:

- <https://www.spec.org/cpu2017/Docs/overview.html>

La imagen siguiente da una idea de esta página.

Benchmarks: good, bad, difficult, and standard

- Q1. What is SPEC?
- Q2. What is a (good) benchmark?
- Q3. What are some common benchmarking mistakes?
- Q4. Should I benchmark my own application?
- Q5. Should I use a standard benchmark?

SPEC CPU 2017 Basics

- Q6. What does SPEC CPU 2017 measure?
- Q7. Should I use CPU 2017? Why or why not?
- Q8. What does SPEC provide?
- Q9. What must I provide?
- Q10. What are the basic steps to run CPU 2017?
- Q11. How long does it take? Is it longer than CPU 2006?

Suites and Benchmarks

- Q12. What is a CPU 2017 "suite"?
- Q13. What are the benchmarks?
- Q14. Are 5nn.benchmark and 6nn.benchmark different?

CPU 2017 Metrics

- Q15. What are "SPECspeed" and "SPECrate" metrics?
- Q16. What are "base" and "peak" metrics?
- Q17. Which SPEC CPU 2017 metric should I use?
- Q18. What is a "reference machine"? Why use one?

Q19. What's new in SPEC CPU 2017?

- a. Primary content
 - 1. Benchmarks and Metrics
 - 2. Source Code: C99, Fortran-2003, C++2003
 - 3. Rules
- b. Power
- c. Documentation
- d. Tools
 - 1. SPECrate smarter setup
 - 2. sysinfo is required
 - 3. Compiler versions are required
 - 4. Image Validation
 - 5. Usability
 - 6. Removed items

Publishing results

- Q20. Where can I find SPEC CPU 2017 results?
- Q21. Can I publish elsewhere? Do the rules still apply?

Transitions

- Q22. What will happen to SPEC CPU 2006?
- Q23. Can I convert CPU 2006 results to CPU 2017?

SPEC CPU 2017 Benchmark Selection

- Q24. What criteria were used?
- Q25. Were some benchmarks 'kept' from CPU 2006?
- Q26. Are the benchmarks comparable to other programs?

Miscellaneous

- Q27. Can I run the benchmarks manually?
- Q28. How do I contact SPEC?
- Q29. What should I do next?

La página anterior resulta muy instructiva para las personas que se acercan por primera vez a este benchmark porque está estructurado en una serie de preguntas y respuestas bastante concretas sobre el diseño y uso de este programa de prueba.

Componentes evaluados del computador

El benchmark SPEC CPU2017 evalúa el rendimiento de tres componentes o partes concretas del sistema informático. ¿De qué tres elementos se trata?

Componente 1	Procesador
Componente 2	Arquitectura de Memoria
Componente 3	Compilador

Resultados del benchmark

El benchmark SPEC CPU2017 expresa el rendimiento del computador mediante dos tipos de medidas (metrics), según se refieran a la velocidad de ejecución (tipo SPECspeed) o a la productividad (tipo SPECrate). ¿Qué mide exactamente cada una de estas dos métricas? ¿Se permite usar OpenMP?

Métrica	¿Que mide?	¿Usa OpenMP?
SPECspeed	Una copia de cada benchmark donde el usuario elige la cantidad de hilos Open MP. Cuántas unidades de tiempo se necesitan para completar un trabajo	SI
SPECrate	Se ejecutan varias copias de cada una de las benchmark de forma concurrente. Cuántos trabajos se pueden completar en una unidad de tiempo.	NO

Suites e índices de rendimiento

Este benchmark evalúa, a grandes rasgos, el rendimiento de tipo aritmético de un computador y lo hace ayudándose de cuatro conjuntos diferentes de programas. Estos conjuntos se denominan suites en inglés. Indique los 4 índices de rendimiento que ofrece el benchmark. ¿Por qué razón se distingue entre aritmética entera y aritmética real?

SPECspeed:	
SPECspeed2017_int_base	SPECspeed2017_fp_base
SPECspeed2017_int_peak	SPECspeed2017_fp_base
SPECrate:	
SPECrate2017_int_base	SPECrate2017_fp_base
SPECrate2017_int_peak	SPECrate2017_fp_base
Se distingue porque las operaciones de aritmética real tienen un coste computacional mayor al de aritmética entera.	

Lenguajes de programación

Indique los lenguajes de programación en que están escritos los programas del benchmark.

C, C++ y Fortran

Criterios de selección de programas

Indique al menos tres criterios empleados por los miembros de la corporación SPEC para seleccionar los programas que integran el benchmark CPU2017.

Criterio 1	Availability of workloads that represent real problems.
Criterio 2	Performance profile: Is the candidate compute bound, spending most of its time in the benchmark source, and little time in IO and system services?

Criterio 3	Portability to a variety of CPU architectures, including ARM, Itanium, Power ISA, SPARC, and x86.
------------	---

Programas de Inteligencia Artificial

Algunos de los programas del benchmark que evalúan la potencia de procesamiento en aritmética entera (suite SPECspeed®2017 Integer) se utilizan en la práctica en la resolución de problemas relacionados con la inteligencia artificial. ¿De qué programas se trata exactamente. Selecciónelos de la siguiente lista.

- 602.gcc
- 631.deepsjeng
- 605.mcf
- 625.x264
- 600.perlbench
- 648.exchange2
- 641.leela

Programas de dinámica de fluidos y molecular

Por otro lado, algunos de los programas del benchmark que evalúan la potencia de procesamiento en aritmética real (suite SPECspeed®2017 Floating Point) se utilizan en la práctica en problemas relacionados tanto con la dinámica de fluidos como con la dinámica molecular. ¿De qué programas específicos se está hablando? Selecciónelos de la siguiente lista.

- 603.bwaves
- 607.cactuBSSN
- 619.lbm
- 621.wrf
- 627.cam4
- 628.pop2
- 638.imagick
- 644.nab
- 649.fotonik3d
- 654.roms

Sistema de referencia

¿Qué máquina se ha usado como sistema de referencia en los cálculos de este benchmark? Indique en la siguiente tabla sus características básicas: año, fabricante, nombre del modelo, procesadores y cores, memoria principal, memoria cache, sistema de almacenamiento, etc. Especifique el máximo de detalle de cada elemento.



Año de producción	2017
Fabricante	Sun Microsystems
Modelo del servidor	Sun Fire V490
Unidad Central de Proceso	2100 MHz UltraSPARC-IV+
Unidad de Memoria	32 GB (32 x 1 GB SDRAM Registered, ECC, 232-pin, Samsung M323S6459ET2-C1LC2)
Unidad de Almacenamiento	300 GB ZFS mirror on 2x 15K RPM 300 GB Fibre Channel drives

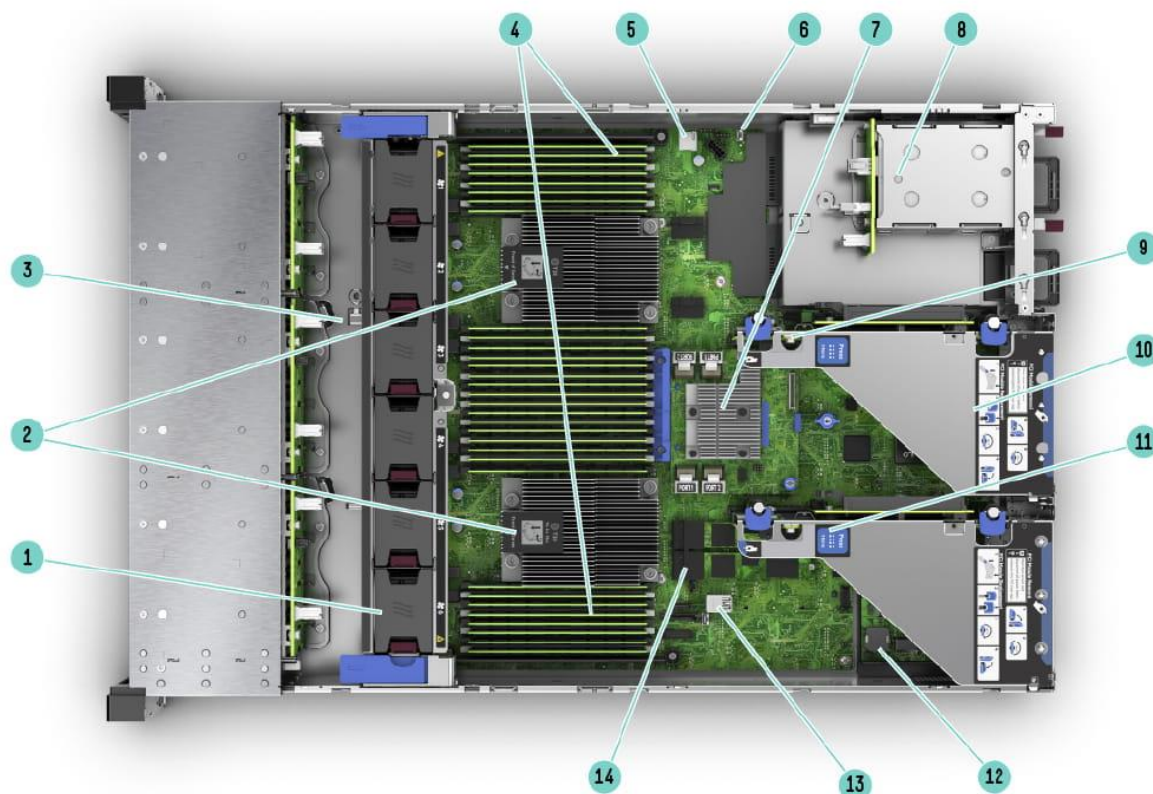
Se puede consultar los resultados de aritmética real (SPECspeed2017_fp_base) de este sistema de referencia en cualquiera de los enlaces siguientes:

- <https://www.spec.org/cpu2017/results/res2017q2/cpu2017-20161026-00002.html>
- <https://www.spec.org/cpu2017/results/res2017q2/cpu2017-20161026-00002.txt>
- <https://www.spec.org/cpu2017/results/res2017q2/cpu2017-20161026-00002.pdf>

Servidor HP ProLiant DL385 Gen10

En este apartado se analiza el rendimiento en aritmética real (coma flotante) del servidor Hewlett Packard Enterprise ProLiant DL385 Gen10, perteneciente a la serie HP ProLiant. Una descripción detallada del producto puede consultarse en la [página web del fabricante](#). También resulta muy interesante [este vídeo](#) en el que podemos ver los distintos elementos que conforman este servidor.





Internal View 8SFF chassis – with optional 2nd CPU, FlexLOM, Smart array shown

- | | |
|--|--|
| 1. Fan cage shown with 6 standard Hot-plug fans (High Performance fans optional) | 2. 2 Processors (heatsinks shown) |
| 3. Optional HPE Smart Storage Battery (not shown) | 4. DDR4 DIMM slots. Shown fully populated in 32 slots (16 per processor) |
| 5. MicroSD card slot (Optional Dual Micro-SD option) | 6. Chassis intrusion detection connector |
| 7. Optional HPE Flexible Smart Array Controller (P408i-a shown) | 8. (Under) Hot Plug redundant HPE Flexible Slot Power supplies |
| 9. Connection for second (optional) riser (Requires second CPU) | 10. Embedded 4x1Gbe NIC |
| 11. Primary PCIe riser, standard (Optional double wide GPU riser) | 12. Optional Flexible LOM slot |
| 13. Internal USB 3.0 connector | 14. Embedded M.2 connectors |

La información con la que se va a trabajar en esta sesión de prácticas corresponde al rendimiento en **coma flotante** de este servidor que está publicada en la web de SPEC en tres formatos distintos:

- <https://www.spec.org/cpu2017/results/res2019q1/cpu2017-20190204-10943.html>
- <https://www.spec.org/cpu2017/results/res2019q1/cpu2017-20190204-10943.txt>
- <https://www.spec.org/cpu2017/results/res2019q1/cpu2017-20190204-10943.pdf>

El formato texto es especialmente útil cuando se quieren copiar valores a una hoja de cálculo, como se verá más adelante. Tómese su tiempo para examinar detenidamente esta información porque será la base para responder a las cuestiones que aparecen a continuación. Tenga en cuenta que la configuración exacta del servidor evaluado es la referida en esta documentación.

Configuración del hardware

Indique las principales características del servidor ProLiant DL385 Gen10: modelo de procesador, número de procesadores, número de núcleos (cores) por procesador, memorias cache incorporadas en el procesador, tamaño de memoria principal y, finalmente, número de discos y tipo.

```
Procesador AMD EPYC 7371
Nº procesadores: 2 chips
Nº núcleos: 32 en cada uno
Memoria cache:
    L1: 64 KB I + 32 KB D on chip per core
    L2: 512 KB I+D on chip per core
    L3: 64 MB I+D on chip per chip, 8 MB shared / 2 cores
Memoria principal: 1 TB (16 x 64 GB 4Rx4 PC4-2666V-L)
Nº discos y tipo: 1 x 400 GB SAS SSD RAID 0
```

Configuración del software

Indique el nombre del sistema operativo del servidor evaluado y los compiladores utilizados para compilar los diferentes programas que integran la suite del benchmark SPEC CPU2017 de aritmética real.

```
SUSE linux Enterprise Server 12 (x86_64) SP3
Kernel 4.4.132-94.33-default

Compiladores:
C/C++: Version 1.2.1 of AOCC
Fortran: Version 4.8.2 of GCC
```

Número de ejecuciones

A la vista de los tiempos de ejecución publicados en la web, ¿cuántas veces se ha ejecutado cada programa del benchmark? De los tiempos de ejecución obtenidos por cada programa, ¿cómo se selecciona el valor que finalmente participa en el cálculo de los índices de rendimiento?

```
3 veces cada uno.

El tiempo elegido es con la MEDIANA.
```

Cálculo de los índices SPECspeed2017_fp

De acuerdo con la información publicada por SPEC, los índices de prestaciones para este servidor son: SPECspeed2017_fp_base = 114 y SPECspeed2017_fp_peak = 119. Compruebe, mediante una hoja de cálculo, que estos índices están correctamente calculados a partir de los tiempos de ejecución de los programas en el servidor y en la máquina de referencia.

Recuerde que la fórmula utilizada para calcular los índices es la media geométrica de las aceleraciones (ratios) obtenidas en la ejecución de los diez programas de la suite. Cada ratio se calcula como el cociente entre el tiempo de ejecución del programa en el sistema de referencia (TR_i) y el tiempo de ejecución del programa en nuestro servidor (T_i):

$$\sqrt[10]{\prod_{i=1}^{10} \left(\frac{TR_i}{T_i} \right)}$$

En Microsoft Excel existe la fórmula MEDIA.GEOM() y en Google esta misma función se denomina GEOMEAN().

Índice	Publicado por SPEC	Calculado
SPECspeed2017_fp_base	114	113,919
SPECspeed2017_fp_peak	119	118,794

Interpretación de SPECspeed2017_fp_base

Explique qué quiere decir que el índice SPECspeed2017_fp_base de este servidor tiene un valor de 114.

Que es 114 veces más rapido

Optimización del compilador

Si nos atenemos a los valores obtenidos de los índices de rendimiento del servidor SPECspeed2017_fp_base y SPECspeed2017_fp_peak, ¿en qué porcentaje ha mejorado el rendimiento del sistema utilizando las opciones de compilación que optimizan la ejecución del benchmark?

11400% para base
11900% para peak

Indique en qué programa de la suite de coma flotante han tenido mayor influencia las opciones de optimización usadas durante el proceso de compilación. Ayúdese de la hoja de cálculo para contestar esta pregunta.

```
638.imagick_s
```

Por la diferencia en segundos de las pruebas entre las versiones

Tiempos de ejecución

¿Qué programa de la suite de coma flotante ha sido el que menos ha tardado en ejecutarse? ¿Y el que más? Tenga cuidado a la hora de hacer la comparación: no es lo mismo comparar tiempos que aceleraciones (es decir, ratios).

```
Menos tiempo 644.nab_s con 78,5 segundos
```

```
Mas tiempo 628.pop2_s con 268 segundos
```

Relación entre SPECspeed2017_fp y SPECspeed2017_int

Con lo visto hasta ahora respecto del rendimiento en aritmética de coma flotante, ¿se atrevería a predecir un valor para el índice de rendimiento SPECspeed2017_int de este servidor? Razone la respuesta.

No porque para la aritmética de coma flotante el juego de instrucciones cambia según el sistema y arquitectura por lo que sus tiempos serán distintos.