

# BENCHMARK (IN COMPUTING)

## Mục đích

- Tìm hiểu về Benchmark trong máy tính
- Biết cách sử dụng và diễn giải các thông số phần cứng được so sánh khi thực hiện Benchmark trên một số phần mềm phổ biến

## Tóm tắt lý thuyết

### Benchmark là gì?

Khi bạn sử dụng công cụ hay chương trình để đánh giá một bộ phận hay toàn bộ hệ thống PC của mình với một độ đo xác định, đem kết quả thu được (đặc trưng bởi từng chương trình đánh giá) so với cơ sở dữ liệu kết quả có sẵn hay với một hệ thống khác thì bạn đã thực hiện công việc “Benchmark”.

Trong lĩnh vực máy tính, Benchmark là việc sử dụng 1 hay nhiều chương trình phần mềm máy tính để tính toán hiệu năng hoạt động của các thiết bị phần cứng dựa trên 1 số đặc tính cụ thể nào đó (CPU clock speed, Memory read/write, Tốc độ tính số chấm động – FPU, Tỷ lệ frame per second khi hiển thị đồ họa...) bằng cách chạy các bài test chuẩn định sẵn, cuối cùng so sánh chúng với các hệ thống khác trên cùng 1 bài test nhằm giúp người sử dụng có 1 cái nhìn tổng quan “sức mạnh” của hệ thống mình, 1 cách tương đối.

Ngoài ra người ta cũng có thể sử dụng thuật ngữ Benchmark để nói về sự tương quan tốc độ và sự ổn định của các phần mềm trên máy tính (Compiler, Database management systems...) hoặc chỉ đơn giản kiểm tra sự đúng đắn của chúng.

Tuy nhiên trong phạm vi bài thực hành này chúng ta chỉ xem xét Benchmark cho phần cứng máy tính

### Tại sao lại cần Benchmark?

Người tiêu dùng luôn mong muốn có được sản phẩm phải thật đáng với đồng tiền họ bỏ ra, tất nhiên không phải người tiêu dùng nào cũng hiểu rõ mình cần gì và cái mình muốn mua có chính xác với nhu cầu chưa. Đối với những sản phẩm liên quan tới PC, may thay chúng ta có được những kết quả Benchmark mà từ đó, người tiêu dùng biết mình nên đầu tư vào sản phẩm nào cho chính xác.

Tuy nhiên cũng xin lưu ý là những kết quả Benchmark chỉ là những giá trị tham khảo chứ không thể dựa hoàn toàn vào đó để kết luận “sức mạnh” nói chung của 1 phần cứng / hệ thống vì mỗi thiết bị thuộc nhiều nhà sản xuất khác nhau có thể có những kiến trúc hoàn toàn khác nhau (Ví dụ: CPU giữa Intel & AMD; GPU giữa nVidia & ATI) dẫn đến có những ưu và khuyết điểm khác nhau. Mặt khác nhà sản xuất có thể chỉnh sửa nhằm tối ưu cho các bài test trên các công cụ Benchmark tuy nhiên chạy thực tế lại không được như thế.

### Benchmark thường quan tâm đến yếu tố gì?

Thông thường, Benchmark là cách so sánh cho những người quan tâm đến lượng thời gian mà họ có, tức là dùng Benchmark để xem thiết bị nào làm việc nhanh hơn với cùng 1 khối lượng công việc (tốc độ, thời gian xử lý).

Bên cạnh đó đôi khi tốc độ chưa phải là tất cả, người ta còn quan tâm đến chất lượng xử lý (vì có xử lý nhanh mà chất lượng tệ hơn thì cũng không phải là tốt). Đây là một trường hợp tiêu biểu của Benchmark trong GPU (Graphic Control Unit), vì ngoài thông số hiển thị frame hình trên 1 giây, người ta còn phải đề cập con số frame/s đó có được khi đang trong ở chế độ nào (Low / Middle / High quality; có khử răng cưa không, nếu có thì bao nhiêu lần...)

### **Phân loại Benchmark**

Như đã đề cập ở trên, trong lĩnh vực Computing, Benchmark thường dùng để đo và so sánh hiệu năng phần cứng và phần mềm (ít hơn). Tuy nhiên dù là phần cứng hay phần mềm thì các phép thử Benchmark đều được áp dụng ở 1 trong 2 trường hợp sau:

- **Benchmark thực tế (Real-world):** Các bài test là những gì mà người dùng sẽ sử dụng hằng ngày, thực sự có nhu cầu khi vận hành trong thực tế.
  - Ví dụ: Benchmark xử lý đồ họa cho một PC game nào đó, Benchmark thời gian chạy khi nén hoặc giải nén file lớn...
- **Benchmark giả lập (Synthetic / Artificial):** Các bài test chỉ là những cuộc thử nghiệm “giả” (rất ít xảy ra trong thực tế) nhằm mục đích tập trung đo một số các thông số xử lý cụ thể nào đó trên thiết bị.
  - Ví dụ: Benchmark thời gian xử lý render hình trên chip GPU khi muốn vẽ một ma trận hình Fractal, Benchmark thời gian tính toán và lưu trữ trên CPU khi nhân, chia nhiều số chấm động rất lớn / rất nhỏ liên tục...

Thông thường người ta thường chạy các Benchmark giả lập mặc dù Benchmark thực tế mới là cái mọi người quan tâm hơn, lý do vì việc tạo ra một môi trường test thực tế phù hợp mục đích của người test luôn phức tạp và khó hơn việc làm “giả” chúng. Với việc dùng Benchmark giả lập chúng ta có thể tạo ra môi trường đồng nhất với mọi bài test cho mọi hệ thống khác nhau và có thể tùy biến các test-case nhằm mục đích đưa ra những kết quả khách quan hơn.

Tuy nhiên điều này cũng đem lại một số khuyết điểm:

- Người tiêu dùng (người review kết quả) không biết liệu cái Benchmark giả lập có bao giờ xảy ra trong thực tế để nó có thể đạt được kết quả đó hay không? Nếu chẳng bao giờ tôi dùng hay gặp chúng, vậy có cần thiết để đo Benchmark chúng? Tuy nhiên do tâm lý của khá nhiều người tiêu dùng là “Thà nó có (mạnh hơn) còn hơn không, mặc dù tôi chẳng bao giờ cần” nên những kết quả Benchmark giả lập vẫn được các nhà sản xuất chú trọng nhiều.
- Tạo ra hệ quả “thiếu lành mạnh”: Người sản xuất chỉ tập trung tối ưu sản phẩm của họ để cho kết quả giả lập thật cao mà quên đi năng lực khi “chiến đấu” thực tế. Ví dụ: Phần mềm Benchmark PCMark 2005 có xu hướng cho điểm các hệ thống dùng CPU của Intel cao hơn AMD, hay 3dMax (Benchmark đồ họa) cũng “thích” chip của ATI hơn nVidia...

### **Nguyên tắc trong Benchmark**

Ở bất kỳ phép đo và so sánh nào, cụ thể trong Benchmark, phải đáp ứng ít nhất 3 nguyên tắc chính sau (3R): *Real-world (thực tế)*, *Relevance (thích hợp)* và *Repeatability (có khả năng lặp lại)*

- **Real-world:** Như đã nói ở trên, người ta luôn muốn việc đo và so sánh cần phải thực hiện trong những môi trường thực tế nhằm có cái nhìn khách quan. Đôi khi có những loại Benchmark mà bạn chẳng biết có bao giờ chúng xảy ra trong thực tế không, có cần thiết với nhu cầu của bạn khi sử dụng máy tính hàng ngày không hay thậm chí thông số đó dùng để

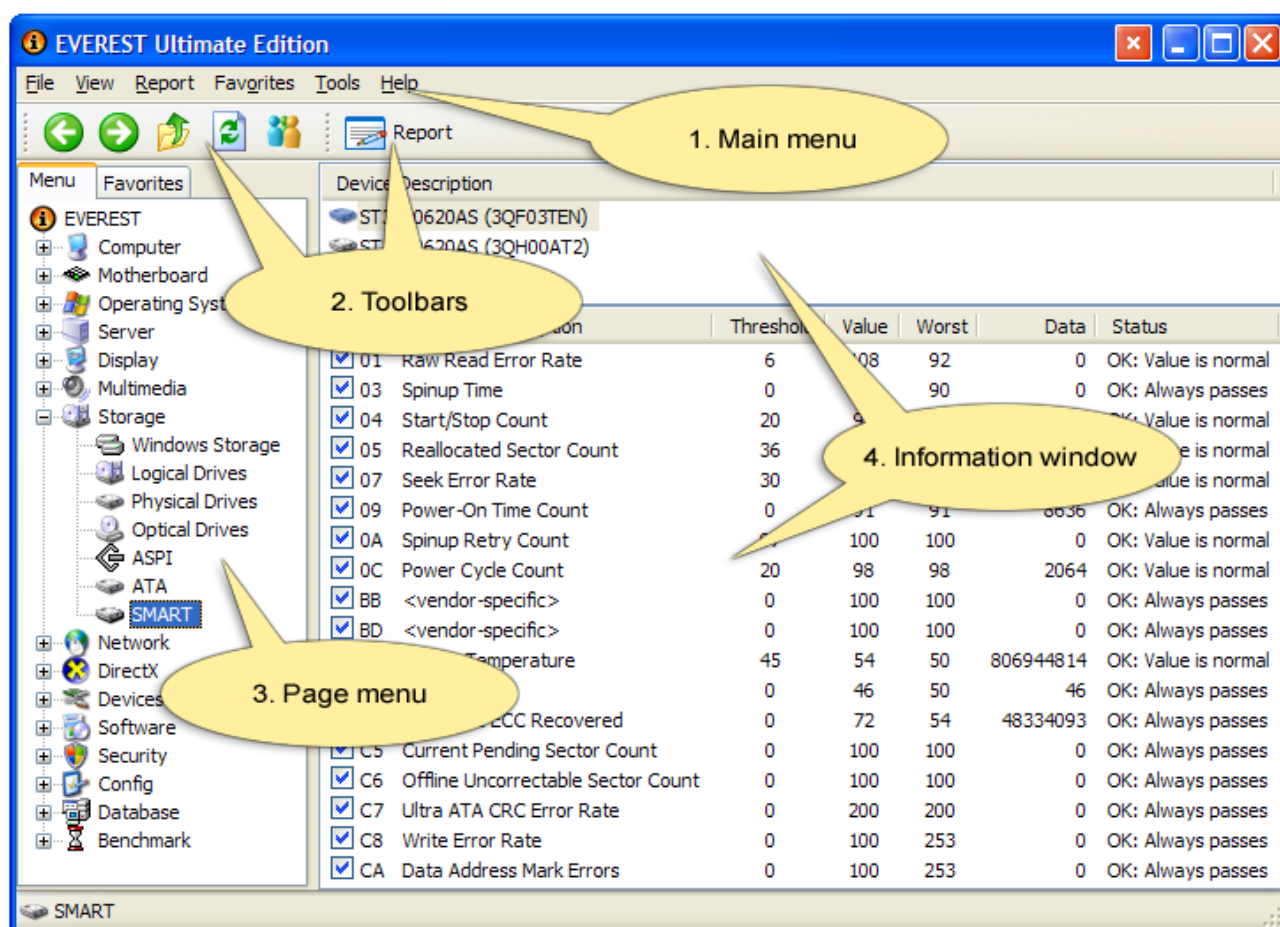
làm gì? Do vậy vấn đề Benchmark trong real-world luôn là mong muốn và nguyên tắc cơ bản trong Benchmark ngày nay.

- **Relevance:** Sẽ là không hợp lý lắm nếu như khi nhìn vào kết quả Benchmark chúng ta kết luận tốc độ của GPU đã ảnh hưởng rất nhiều đến tốc độ xử lý của phần mềm Microsoft Excel, hoặc ta muốn Benchmark hệ thống để so sánh tốc độ xử lý của phần mềm Notepad. Do vậy Benchmark sẽ chẳng có ý nghĩa gì nếu chúng không liên quan đến nhu cầu và ý nghĩa thực tế của thiết bị được Benchmark.
- **Repeatability:** Bạn có 1 thiết bị phần cứng cao cấp với rất nhiều tính năng mới nhất tuy nhiên bạn không thể dựa vào đó để kết luận ngay nó sẽ nhanh hơn rất nhiều so với các thiết bị phần cứng khác mà không được trang bị những tính năng đó, vì có thể những thiết bị kia sẽ có những tính năng ngược lại mà thiết bị của bạn không có. Nói cách khác, kết quả Benchmark phải có giá trị so sánh cụ thể bằng số. Ngoài ra giữa các lần đo Benchmark lặp lại, sự biến thiên sai số không được quá lớn (cỡ vài chục %), nhưng vài % là có thể chấp nhận được (do sự biến thiên thay đổi liên tục của môi trường test).

### **Công cụ Everest**

Hiện nay trên thế giới có rất nhiều các công cụ tổng quát chuyên dành cho Benchmark PC như Everest, WorldBench, PCMark, Sisoft Sandra, HWinfo32... hay các công cụ Benchmark dành cụ thể cho 1 chức năng nào đó, ví dụ xử lý đồ họa như 3dMax. Ở phạm vi bài hướng dẫn thực hành này xin đề cập đến công cụ đơn giản và khá hiệu quả là Everest nhằm giúp sinh viên có cái nhìn cụ thể của một công cụ Benchmark.

Everest là một phần mềm quản lý thông tin tổng quát (phần cứng & phần mềm) trong PC, bên cạnh đó nó còn cung cấp các bài test nhằm Benchmark các hệ thống khác nhau. Lưu ý trong Everest chỉ sử dụng Benchmark giả lập (Synthetic) nhằm giúp người dùng có cái nhìn nhanh chóng và dễ dàng so sánh với các hệ thống cùng loại khác chứ không có các Benchmark thực tế (Real-world). Cuối cùng Everest còn cung cấp các công cụ test và chuẩn đoán đĩa cứng, bộ nhớ và màn hình hay sự ổn định (nhiệt độ, tần số hoạt động) của hệ thống xem có hoạt động tốt không (chứ không có mục đích so sánh như trong Benchmark)



Giao diện chính của Everest

Quản lý một số thông tin phần cứng quan trọng:

The screenshot shows the EVEREST Ultimate Edition interface. The left sidebar has a tree view with categories like Computer, Motherboard, CPU, Memory, Chipset, BIOS, ACPI, Operating System, etc. The 'CPU' category is selected. The main window displays a table of CPU properties.

Field	Value
<b>CPU Properties</b>	
CPU Type	Mobile DualCore Intel Core Duo T2300, 1666 MHz (10 x 167)
CPU Alias	Yonah
CPU Stepping	C0
Instruction Set	x86, MMX, SSE, SSE2, SSE3
Original Clock	1667 MHz
Min / Max CPU Multiplier	6x / 10x
Engineering Sample	No
L1 Code Cache	32 KB per core
L1 Data Cache	32 KB per core
L2 Cache	2 MB (On-Die, ECC, Full-Speed)
<b>Multi CPU</b>	
CPU #1	IntelGenuine Intel(R) CPU T2300 @ 1.66GHz, 1663 MHz
CPU #2	IntelGenuine Intel(R) CPU T2300 @ 1.66GHz, 1663 MHz
<b>CPU Physical Info</b>	
Package Type	479 Ball uFCBGA / 478 Pin uFCPGA
Package Size	3.5 cm x 3.5 cm
Transistors	151.6 million
Process Technology	8M, 65 nm, CMOS, Cu, Low-K Inter-Layer, 2nd Gen Strained Si
Die Size	90 mm2
Core Voltage	0.950 - 1.262 V
I/O Voltage	0.950 - 1.262 V
<b>CPU Manufacturer</b>	
Company Name	Intel Corporation
Product Information	<a href="http://www.intel.com/products/processor">http://www.intel.com/products/processor</a>

At the bottom of the window, a status bar shows: BDLL: 2.4.273.0, CPU: 1663 MHz, TSC: 1663 MHz, T: 13.62 sec, FT: 0.38 sec.

Thông tin tổng quát CPU

The screenshot shows the EVEREST Ultimate Edition interface with the 'CPUID' category selected in the left sidebar. The main window displays a table of CPUID features.

Field	Value
<b>CPUID Features</b>	
<input type="checkbox"/> 1 GB Page Size	Not Supported
<input type="checkbox"/> 36-bit Page Size Extension	Not Supported
<input type="checkbox"/> Address Region Registers (ARR)	Not Supported
<input type="checkbox"/> CPL Qualified Debug Store	Not Supported
<input checked="" type="checkbox"/> Debug Trace Store	Supported
<input checked="" type="checkbox"/> Debugging Extension	Supported
<input type="checkbox"/> Direct Cache Access	Not Supported
<input type="checkbox"/> Dynamic Acceleration Technology (IDA)	Not Supported
<input checked="" type="checkbox"/> Fast Save & Restore	Supported
<input type="checkbox"/> Hyper-Threading Technology (HTT)	Not Supported
<input checked="" type="checkbox"/> Invariant Time Stamp Counter	Supported
<input type="checkbox"/> L1 Context ID	Not Supported
<input checked="" type="checkbox"/> Local APIC On Chip	Supported
<input checked="" type="checkbox"/> Machine Check Architecture (MCA)	Supported
<input checked="" type="checkbox"/> Machine Check Exception (MCE)	Supported
<input type="checkbox"/> Memory Configuration Registers (MCR)	Not Supported
<input checked="" type="checkbox"/> Memory Type Range Registers (MTRR)	Supported
<input checked="" type="checkbox"/> Model Specific Registers (MSR)	Supported
<input type="checkbox"/> Nested Paging	Not Supported
<input checked="" type="checkbox"/> Page Attribute Table (PAT)	Supported
<input checked="" type="checkbox"/> Page Global Extension	Supported
<input checked="" type="checkbox"/> Page Size Extension (PSE)	Supported
<input checked="" type="checkbox"/> Pending Break Event	Supported
<input checked="" type="checkbox"/> Physical Address Extension (PAE)	Supported
<input type="checkbox"/> Safer Mode Extensions (SMX)	Not Supported
<input type="checkbox"/> Secure Virtual Machine Extensions (Pacifica)	Not Supported
<input checked="" type="checkbox"/> Self-Snoop	Supported
<input checked="" type="checkbox"/> Time Stamp Counter (TSC)	Supported

At the bottom of the window, a status bar shows: BDLL: 2.4.273.0, CPU: 1663 MHz, TSC: 1663 MHz, T: 13.62 sec, FT: 0.38 sec.

Công nghệ CPU hỗ trợ

The screenshot shows the EVEREST Ultimate Edition interface. The left sidebar has 'Motherboard' selected under the 'Computer' category. The main window displays a table of motherboard properties.

Field	Value
<b>Motherboard Properties</b>	
Motherboard ID	<DMI>
Motherboard ID (Quanta)	MA1 01234567
Motherboard Name	Hewlett-Packard Presario V2000 (EV849PA#AKL)
<b>Front Side Bus Properties</b>	
Bus Type	Intel AGTL+
Bus Width	64-bit
Real Clock	167 MHz (QDR)
Effective Clock	666 MHz
Bandwidth	5331 MB/s
<b>Memory Bus Properties</b>	
Bus Type	Dual DDR2 SDRAM
Bus Width	128-bit
DRAM:FSB Ratio	8:5
Real Clock	267 MHz (DDR)
Effective Clock	533 MHz
Bandwidth	8530 MB/s
<b>Chipset Bus Properties</b>	
Bus Type	Intel Direct Media Interface

At the bottom, a status bar shows: BDLL: 2.4.273.0, CPU: 1663 MHz, TSC: 1663 MHz, T: 13.62 sec, FT: 0.38 sec.

### Motherboard

The screenshot shows the EVEREST Ultimate Edition interface with 'Memory' selected in the left sidebar. The main window displays a table of memory module properties.

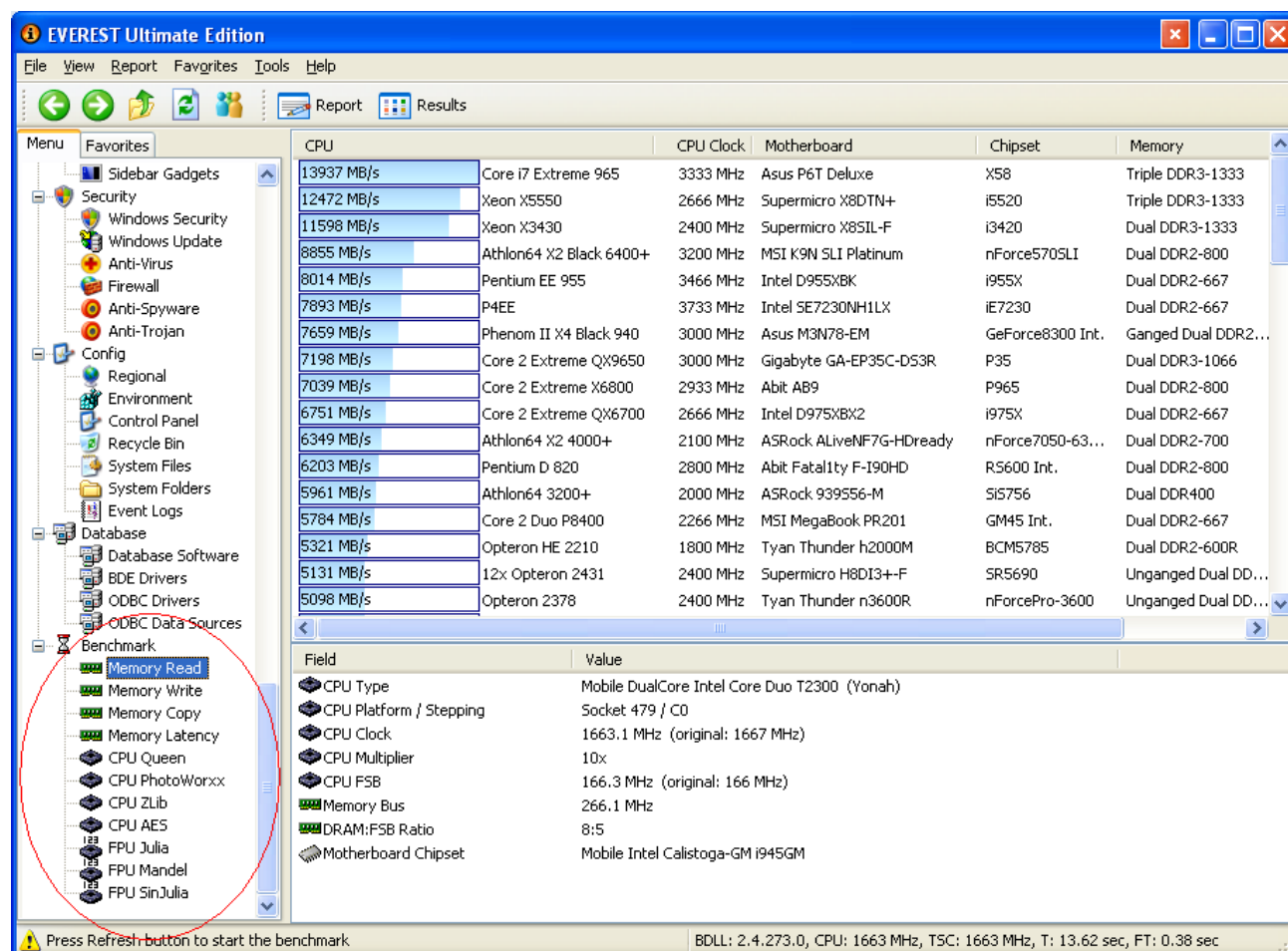
Field	Value
<b>Device Description</b>	
DIMM1: Samsung M4 70T6554CZ3-CD5	
DIMM3: Infineon 64T64020HDL3.7A	
<b>Memory Module Properties</b>	
Module Name	Samsung M4 70T6554CZ3-CD5
Serial Number	08093960h (1614350600)
Manufacture Date	Week 38 / 2005
Module Size	512 MB (2 ranks, 4 banks)
Module Type	SO-DIMM
Memory Type	DDR2 SDRAM
Memory Speed	DDR2-533 (266 MHz)
Module Width	64 bit
Module Voltage	SS1L 1.8
Error Detection Method	None
Refresh Rate	Reduced (7.8 us), Self-Refresh
<b>Memory Timings</b>	
@ 266 MHz	5-4-4-11 (CL-RCD-RP-RAS) / 15-28-3-4-2-2 (RC-RFC-RRD-WR-WTR-RTP)
@ 266 MHz	4-4-4-11 (CL-RCD-RP-RAS) / 15-28-3-4-2-2 (RC-RFC-RRD-WR-WTR-RTP)
@ 200 MHz	3-3-3-8 (CL-RCD-RP-RAS) / 11-21-2-3-2-2 (RC-RFC-RRD-WR-WTR-RTP)
<b>Memory Module Features</b>	
Analysis Probe	Not Present
FET Switch External	Disabled
Weak Driver	Supported
<b>Memory Module Manufacturer</b>	
Company Name	Samsung
Product Information	<a href="http://www.samsung.com/global/business/semiconductor">http://www.samsung.com/global/business/semiconductor</a>

At the bottom, a status bar shows: BDLL: 2.4.273.0, CPU: 1663 MHz, TSC: 1663 MHz, T: 13.62 sec, FT: 0.38 sec.

### Bộ nhớ RAM

**Benchmark:**

Everest hỗ trợ 3 Benchmark chính dành cho: Memory, CPU, FPU (Floating Point Unit) được chia ra nhiều test-case:



Cửa sổ danh sách phía trên phải là những thông số của những hệ thống trong cơ sở dữ liệu trước đó đã được đo.

**Ý nghĩa các test-case trong Everest:**

- **Memory Read:** Benchmark này đo bằng thông tối đa đạt được khi thực hiện thao tác đọc trên bộ nhớ RAM, mã nguồn test-case được viết bằng hợp ngữ nên tối ưu cho nhiều hệ thống xử lý CPU khác nhau. Test-case này sẽ đọc một khối 16MB bộ nhớ, với data buffer là 1 MB từ bộ nhớ hệ thống đưa vào CPU theo 1 chiều thẳng liên tục không ngừng (no break). Lưu ý là để tránh việc những hệ thống có hỗ trợ con-current threads sẽ giành được lợi thế lớn, Benchmark này chỉ sử dụng 1 processor và 1 thread để xử lý thôi.
- **Memory Write:** Giống như Memory Read, ghi từ CPU sang bộ nhớ hệ thống
- **Memory Copy:** Giống như Memory Read, copy khối 8MB bộ nhớ, 1 MB data buffer từ chỗ này sang chỗ khác thông qua CPU
- **Memory Latency:** Đo độ trễ thông thường khi CPU đọc dữ liệu từ bộ nhớ chính. Memory Latency có nghĩa là thời gian trễ tính từ lúc CPU bắt đầu thực hiện lệnh đọc từ bộ nhớ chính cho đến khi dữ liệu thực sự từ bộ nhớ chính đã nằm trong thanh ghi bên trong CPU. Khối dữ liệu để đọc là 1 MB với độ trượt là 1KB.

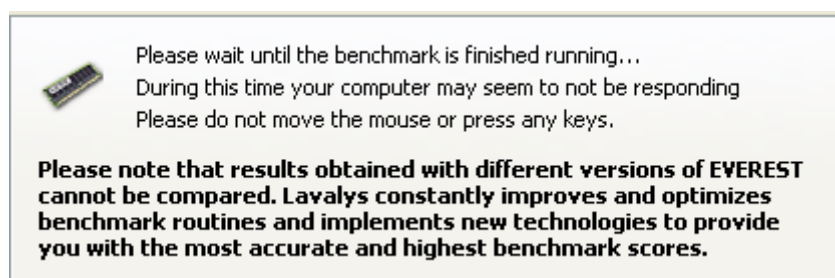


- **CPU Queen:** Đây là một benchmark đơn giản của việc dự đoán nước đi trong bài toán “Queen’s problem” (cờ vua) <http://mathworld.wolfram.com/QueensProblem.html> nhằm đánh giá khả năng tính toán của CPU, trong phép thử này chỉ dùng khoảng ít hơn 1MB bộ nhớ.
- **CPU PhotoWorxx:** Benchmark này thực hiện nhiều tasks trong quá trình xử lý ảnh số của một bức ảnh hệ màu RGB rất lớn:
  - Fill
  - Flip
  - Rotate90R (rotate 90 degrees CW)
  - Rotate90L (rotate 90 degrees CCW)
  - Random (fill the image with random coloured pixels)
  - RGB2BW (colour to black & white conversion)
  - Difference
  - Crop

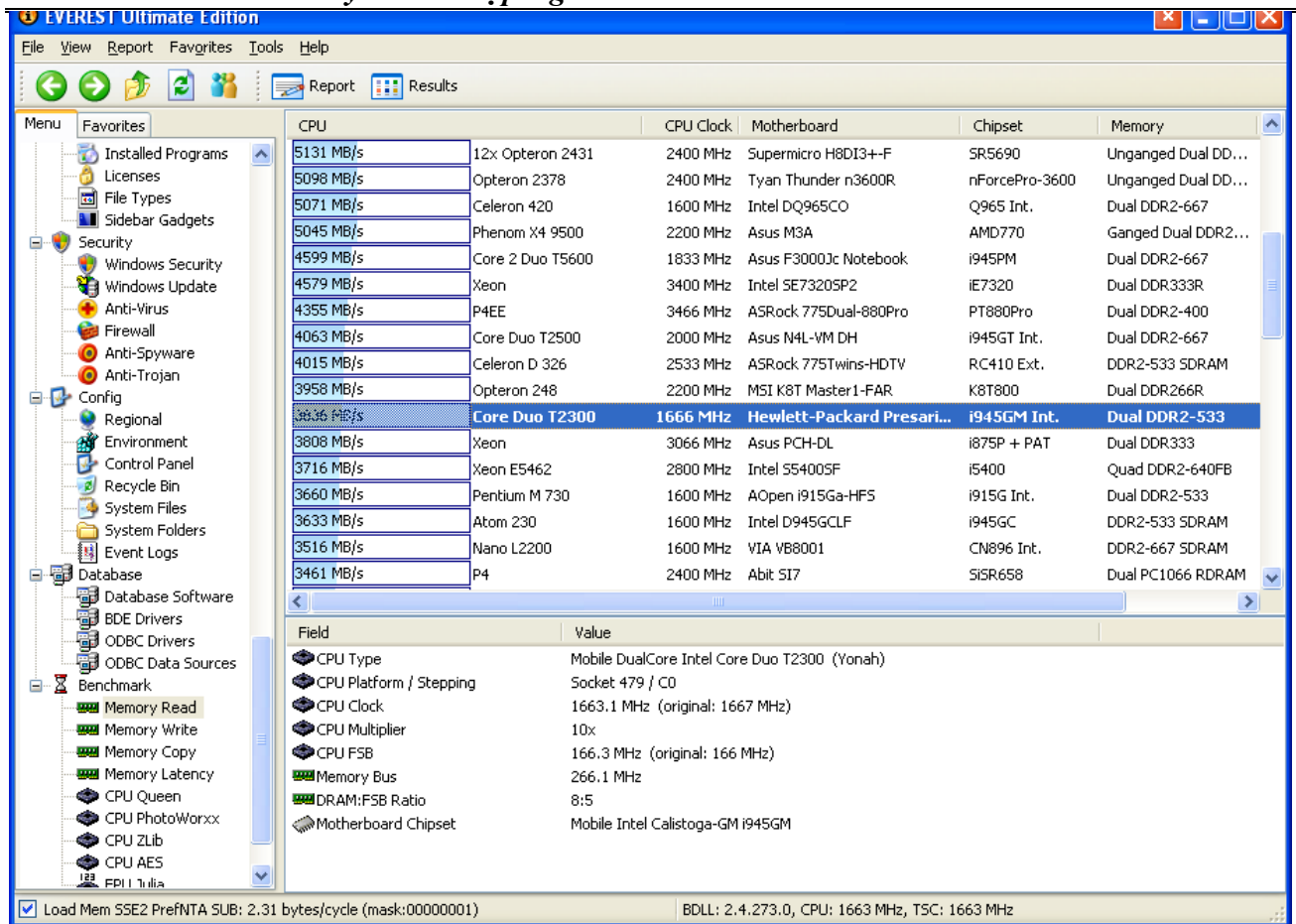
Phép thử này nhằm làm cho thành phần tính toán số học của CPU phải làm việc cật lực. Bởi vì phép thử này được tiến hành trong điều kiện băng thông bộ nhớ lớn nên sẽ không hiệu quả nếu như dùng multi-thread để xử lý do vậy CPU PhotoWorxx chỉ dùng 1 thread với mã lệnh hợp ngữ trên x86.

- **CPU ZLib:** Benchmark này thực hiện nén file lớn theo chuẩn ZLib version 1.2.3 (<http://www.zlib.net>) nhằm đánh giá khả năng phối hợp giữa CPU và bộ nhớ. CPU ZLib chỉ sử dụng các mã lệnh hợp ngữ x86, một nhân, một thread.
- **CPU AES:** Tương tự như CPU ZLib, CPU AES sẽ mã hóa và giải nén file theo chuẩn AES 256 bit (còn gọi là Rijndael), Benchmark này được viết bằng mã hợp ngữ x86, một nhân, một thread, dùng 48MB bộ nhớ.
- **FPU Julia:** Benchmark này sẽ tính toán trên hệ số dấu chấm động đơn (32 bits) trong bài toán lớn vẽ hình Julia fractal, Benchmark này được viết bằng mã hợp ngữ x86, một nhân, một thread, dùng 1MB bộ nhớ.
- **FPU Mandel:** Benchmark này sẽ tính toán trên hệ số dấu chấm động kép (64 bits) trong bài toán lớn vẽ hình Mandelbrot fractal, Benchmark này được viết bằng mã hợp ngữ x86, một nhân, một thread, dùng 1MB bộ nhớ.
- **FPU SinJulia:** Benchmark này sẽ tính toán trên hệ số dấu chấm động mở rộng (80 bits) trong bài toán lớn vẽ hình Julia fractal được chỉnh sửa, Benchmark này được viết bằng mã hợp ngữ x86, một nhân, một thread, dùng 1MB bộ nhớ.

Nếu bạn muốn Benchmark hệ thống của bạn tại 1 test-case cụ thể (Ví dụ Memory Read) thì nhấn nút Refresh ở trên thanh Toolbar hay nhấn F5 để Everest bắt đầu đo.





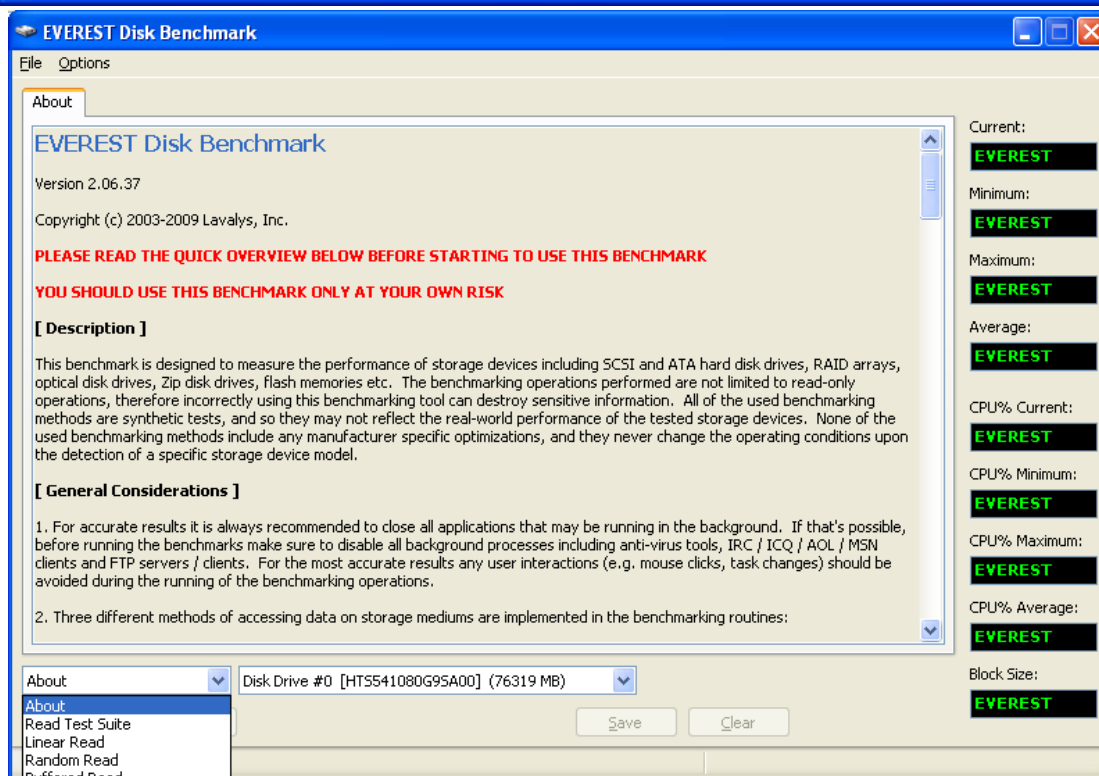
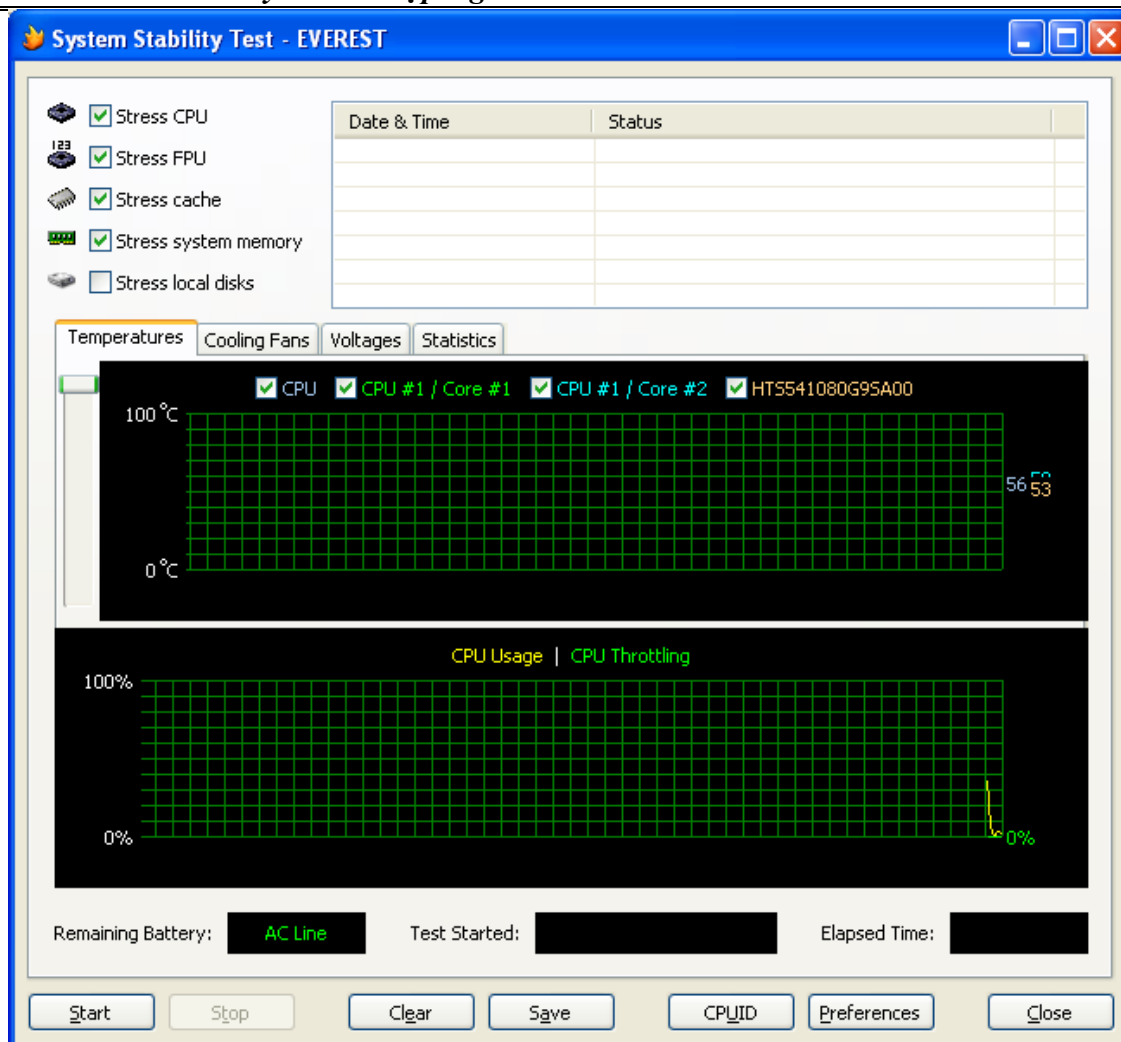


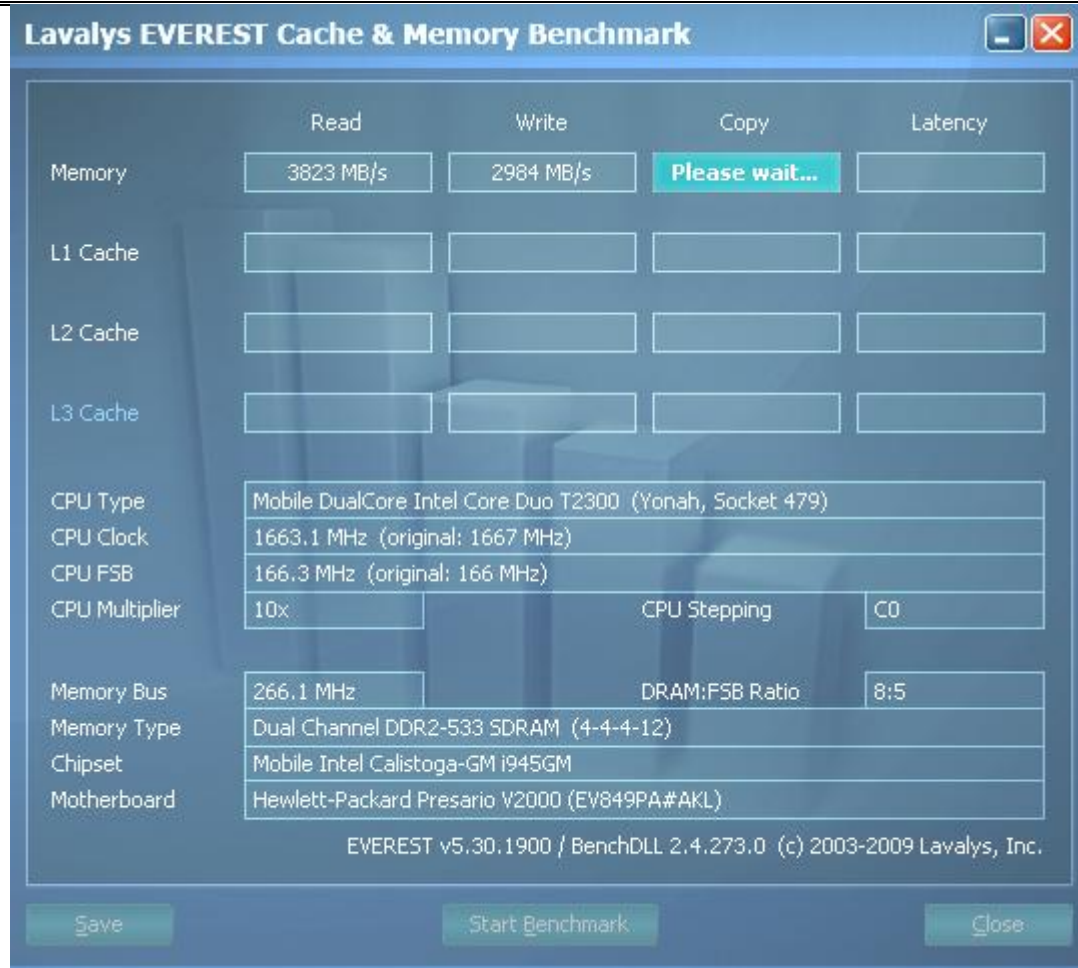
Kết quả Benchmark hệ thống sau khi đo với phép thử Memory Read

### Lưu ý:

- Để có thể đạt được kết quả cao nhất có thể, trước khi Benchmark nên tắt tất cả các ứng dụng đang chạy, chỉ giữ lại những gì tối thiểu nhất có thể (Ví dụ chỉ giữ lại phần mềm Everest).
- Ngoại trừ test-case Memory Latency (độ trễ của bộ nhớ), điểm càng cao có nghĩa là càng tốt
- Đối với Memory Read và Memory Write test-case: Đơn vị đo lường là MB (Megabytes)/s
- Memory Latency: ns (nanosecond =  $10^{-9}$  s)
- CPU ZLib: KB (Kilobytes) / s
- Còn lại: Điểm số tương đối do test-case đặt ra

Ngoài ra còn có một chức năng khác để đo và chuẩn đoán hoạt động của hệ thống, ổ cứng, hay bộ nhớ (Trong phần Tools)...





## Tài liệu tham khảo

1. [http://en.wikipedia.org/wiki/Benchmark\\_%28computing%29](http://en.wikipedia.org/wiki/Benchmark_%28computing%29)
2. Google (Benchmark Computing keyword)

## Bài tập

Hãy tìm hiểu các công cụ PC Benchmark khác như PCMark, WorldBench và so sánh với Everest, đặc biệt nhấn mạnh với các test-case mà chúng đã dùng để đo hiệu năng.

## Mở rộng

Ngoài PC Benchmark còn có những phần mềm chuyên để đo và so sánh hiệu năng của 1 chức năng / thành phần cụ thể trên PC (Ví dụ như đồ họa hay độ nhạy chuột dpi). Hãy tìm hiểu xem đó là những phần mềm nào và cho biết chúng dùng phép thử gì để đo?