Benchmark suites to measure computer performance

ICCA'04

Departamento de Informática
Universidade do Minho – 22/01/04

Hernâni Correia

Outline

- Motivation
- Benchmarking overview
- Generic benchmarks
- TPC benchmarks
- SPEC benchmarks
- SPEC CPU2000 benchmark
- Conclusions

Motivation

By nature everyone wants the best and the fastest machine …

- What is meant with "fastest machine?"
- How is "speed" measured?
- The same machine is the fastest under all circumstances?

Benchmarking overview

Comparing different design alternatives:

- Speed

$$\frac{ExecutionTimeY}{ExecutionTimeX} = n$$

- Performa

Key measurement is time !

$$n = \frac{ExecutionTimeY}{ExecutionTimeX} = \frac{\frac{1}{PerformanceY}}{\frac{1}{PerformanceX}} = \frac{PerformanceX}{PerformanceY}$$

Benchmarking overview

Measuring execution time

- ☺ Processor speed
- 😐 Disk and memory accesses
- 😐 Amount and type of memory
- 😐 Operating system
- ☹ Compilers
- ☹ Level of optimization
- ☹ …

Benchmarking overview

Benchmarking can be defined has a task of measuring the performance of a system/subsystem (or application) on a task or on a well defined set of tasks.

The task/workload that is measured is the so called benchmark.

The user's own workload on it's own system is the best benchmark!

Benchmarking overview

Benchmarking: what's important?

- The precise hardware configuration: type of processor, clock speed, number of CPUs, memory size, cache size, video processor and memory, bus, disk speed, and so on.
- The operating system environment like OS version, filesystem, number of concurrent users, etc.
- The version of the benchmark used.
- The program language used (same program could have different execution times if implemented in different languages).
- Compiler used and optimizing level during compilation of benchmarks.

Benchmarking overview

Types of benchmarks:

- ***Real programs***
- ***Kernels***
- ***Toy benchmarks***
- ***Synthetic benchmarks***

Generic benchmarks

- MIPS

  MIPS (or Million Instructions per second) has been one alternative to metrics that use only time.
  MIPS specify the instruction execution rate but it's dependent on the instruction set, making difficult to compare MIPS of different machines with different instruction sets.

- Whetstone and Dhrystone benchmarks

  Whetstone benchmark was the first intentionally written to measure computer performance and was designed to simulate floating point numerical applications.
  Dhrystone is a benchmark program written for testing system's integer performance.

Generic benchmarks

- Lynpack benchmark
  - Derived from a real application which was originated as a collection of linear algebra subroutines in Fortran. It tests **floating point** performance and results are presented in Mflops (millions of floating point instructions per second).
  - It provides three separate benchmarks: Fortran n = 100, Linpack n=1000, HPL (Highly Parallel Computing).
  - HPL is a software package that generates and solves a random dense linear system of equations on distributed-memory computers using 64-bit floating point arithmetic.
  - HPL is the benchmark used for the Top500 report.

TPC benchmarks

- TPC (Transaction Processing Performance Council) - consortium of vendors that defines benchmarks for transaction processing and database domains.
- Define how the tests should run, how system price should be measured and how the results should be reported.

TPC benchmarks

- TPC-D: database benchmark, whose intent is to simulate ad hoc queries characterizing the performance of decision support systems.
- TPC-W is a transactional web benchmark.
  - Workload on the system is performed in a controlled internet commerce environment.
  - Simulates the activities of a business oriented transactional web server.

SPEC benchmarks

- SPEC (Standard Performance Evaluation Corporation) - nonprofit consortium made up by hardware/software vendors, universities, customers and different consultants.
- Intention: develop technically credible and objective system-level benchmarks.
- Benchmarks are derived from real programs
  - placing on the system real workloads.
  - producing realistic results.

SPEC benchmarks

- **SPECmail2001** – is a mail server benchmark (based on Internet SMTP and POP3 protocols).
- **SPECweb99** – is the benchmark for measuring performance of web servers.
- **SPECjvm98** – measures performance of Java Virtual Machines

SPEC CPU2000 benchmark

- CPU2000 benchmark replace CPU95 in measuring the performance of the computer's processor (CPU), memory architecture and compilers on the tested system.
- Reference machine: Sun Ultra10 workstation with a 300 MHz SPARC processor and 256 MB of memory.
- It has two different components: CINT2000 and CFP2000.

SPEC CPU2000 benchmark

CINT2000

Benchmark	Language	Resident size (Mb)	Virtual size (Mb)	Description
164.gzip	C	181	200	Compression
175.vpr	C	50	55.2	FPGA circuit placement and routing
176.gcc	C	155	158	C programming language compiler
181.mcf	C	190	192	Combinatorial optimization
186.crafty	C	2.1	4.2	Game playing: Chess
197.parser	C	37	62.5	Word processing
252.eon	C++	0.7	3.3	Computer visualization
253.perlbmk	C	146	159	Perl programming language
254.gap	C	193	196	Group theory, interpreter
255.vortex	C	72	81	Object-oriented database
256.bzip2	C	185	200	Compression
300.twolf	C	1.9	4.1	Place and route simulator

SPEC CPU2000 benchmark

CFP2000

Benchmark	Language	Resident size (Mb)	Virtual size (Mb)	Description
168.wupwise	F77	176	177	Physics: Quantum chromodynamics
171.swim	F77	191	192	Shallow water modelling
172.mgrid	F77	56	56.7	Multigrid solver: 3D potential field
173.applu	F77	181	191	Partial differential equations
177.mesa	C	9.5	24.7	3D graphics library
178.galgel	F90	63	155	Computational fluid dynamics
179.art	C	3.7	5.9	Image recognition/neural networks
183.equake	C	49	51.1	Seismic wave propagation simulation
187.facerec	F90	16	18.5	Image processing: Face recognition
188.ammp	C	26	30	Computational chemistry
189.lucas	F90	142	143	Number theory/primarily testing
191.fma3d	F90	103	105	Finite-element crash simulation
200.sixtrack	F77	26	59.8	Nuclear physics accelerator design
301.apsi	F77	191	192	Meteorology: Pollutant distribution

Benchmark suites to measure computer performance

SPEC CPU2000 benchmark

- **Metrics**

speed / normalized ratio
- **SPECint2000 / SPECfp2000**
- **SPECint_base2000 / SPECfp_base2000**

normalized throughput
- **SPECint_rate2000 / SPECfp_rate2000**
- **SPECint_rate_base2000 / SPECfp_rate_base2000**