

## Benchmark програми

### 1. Същност на Benchmark програмите.

Терминът **benchmark** на английски означава нещо, което се използва като стандарт, отправна точка, спрямо което се оценяват или измерват други неща. Терминът произлиза от гравираниите хоризонтални вдлъбнатини, които земемерите поставяли в скални структури, за да се поставят в тях нивелирни пръти, с които се отчитала височината.

В компютърните науки, бенчмаркът е действие, при което се стартира компютърна програма или набор програми, за да се получи сравнение на характеристиките на даден обект с някакъв еталон, като се използва набор от стандартни тестове.

Когато се говори за използване на бенчмарк при хардуера, се разбира **сравняване на производителността на две системи**. В българския език бенчмарк тестът обикновено се превежда като **еталонен тест** или **тест за производителност**, а бенчмарк програмата като **програма за еталонно тестване**.

### 2. Необходимост от еталонен софтуер. Предназначение.

С развитието на компютърните архитектури става все по-трудно да се сравнява производителността на различни компютърни системи, просто като се гледат техните спецификации. Например, процесорите Pentium 4 по принцип работят на по-висока тактова честота, отколкото процесорите Athlon XP, което не означава непременно, че имат по-голяма изчислителна мощ. Това е така, защото те имат различна архитектура и различен начин за изпълнение на изчисленията. С други думи необходимо е да се намерят такива критерии за сравнение, които да не зависят от архитектурата на системата, а само от постигнатите резултати.

### 3. Типове еталонен софтуер.

Съществуват 4 основни типа, на които се делят еталонните тестове:

- **Тестове на основата на приложения** (application-based test). Те стартират реални програми и измерват времето за тяхното изпълнение. При това се възпроизвежда максималната активност на приложенията, за създаване на най-голямо натоварване на системата. Те са най-доброто решение за определяне на производителността на цялата система в реални условия. Представители: ZD Winstone и, частично, 3DMark.
- **Възпроизвеждащи тестове** (playback test). Те използват протоколите на системните извиквания, направени по време на определени действия на програмата (например, графически извиквания или използване на диска), а след това ги възпроизвеждат отделно (в изолация). Дават най-реални резултати при тестването на отделни подсистеми, като твърдия диск, CD-ROM, графическата подсистема и др. Представители: CD Winbench и др.
- **Искуствени тестове** (synthetic test). Имитират активността на програмите в отделни подсистеми на компютъра. Най-често се използват при тестване производителността на процесора. Представители: CliBench MK III SMP, CPUMark
- **Контролни тестове** (inspection test). Не се опитват да имитират активността на приложенията, а вместо това се обръщат директно към отделните подсистеми. Най-подходящи са при тестване с цел откриване на грешки и за стабилност на работата на системата.

На резултатите от тестовете не трябва да се вярва безрезервно. Например, някои производители на хардуер специално нагласяват своите драйвери към изкуствените тестове. Освен това един и същи хардуер може да показва много добри резултати с определени тестове, а с други – по-лоши, и обикновено производителите дават резултатите само от тестовете, които представят продуктите им в най-добрата светлина.

Все пак бенчмарковете си остават най-лесния начин за тестване на системата и следващо сравняване с резултатите от тестване на други системи.

#### 4. Характеристики и мерни единици, използвани за измерване на производителността.

Програмите за еталонни тестове използват следните параметри и мерни единици за измерване на производителността:

- **Access time** – време за достъп. Време, което е необходимо на дадено устройство да намери определена информация и да я направи достъпна за обработване.
- **Average seek time** – средно време за достъп до случаен сектор на диска. Характеризира реалната скорост на търсене на твърдия диск или CD/DDVD при четене на фрагментирани файлове и при паралелно използване на диска от няколко приложения.
- **Average linear speed** – средна скорост на линейно четене/запис на данни. Отразява производителността на мултимедийните приложения при четене/запис на големи непрекъснати файлове и запис на цифрово видео в реално време.
- **Dhrystone** – програма, тестваща производителността на системата при работа с цели числа. Тестът се ограничава само в централния процесор и не изпълнява функциите на вход/изход или извиквания на операционната система. Dhrystones per second - колко пъти програмата може да бъде изпълнена за 1 секунда.
- **Whetstone** - програма, тестваща производителността на системата при работа с дробни числа (с плаваща запетая). Тестът се ограничава само в централния процесор и не изпълнява функциите на вход/изход или извиквания на операционната система.
- **FLOPS** - съкращение от floating-point operations per second (операции с плаваща запетая за 1 секунда). Обикновено се използва за оценка на производителността на процесора. MFLOPS - един милион операции с плаваща запетая за секунда. GFLOPS - един милиард операции с плаваща запетая за секунда.
- **MIPS** - съкращение от million instructions per second (милион инструкции за секунда). Една от най-старите мерни единици за скорост на компютъра. Измерва количеството машинни инструкции, които компютърът е способен да извърши за 1 секунда. Това е неточна мярка, тъй като различните инструкции могат да отнемат различно време за изпълнение. Освен това, MIPS зависи само от процесора, докато скоростта на реалните приложения често се ограничава от много други фактори. Поради това все по-рядко се използва в тестовете за производителност
- **FPS** - съкращение от Frames Per Second (кадри за секунда). Характеризира скоростта за работа на системата с графика и видео.
- **Random access/read speed** – скорост за достъп/четене на случайно разпръснати по целия диск блокове, с размер от 0,5 KB до 64 KB.
- **DTR** - data transfer rate - скорост на прехвърляне на данни

## **5. Организации с идеална цел, разработващи софтуер за тестване на производителността, в качеството на индустриални стандарти**

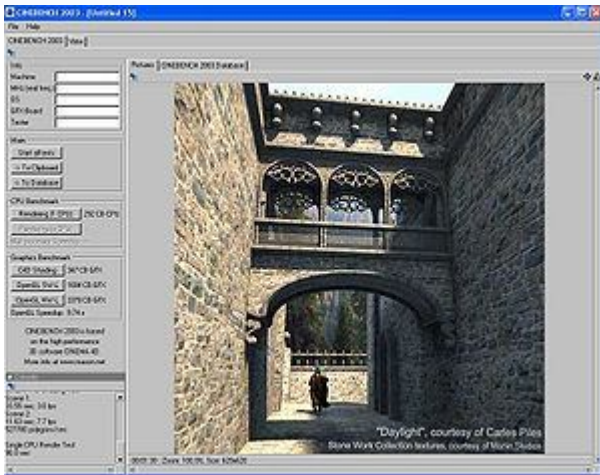
- **ВАРСО** - Business Applications Performance Corporation – създава бенчмаркове за персонални компютри, базирани на популярни софтуерни приложения и операционни системи. Продукти – SYSmark, MobileMark, WebMark
- **ЕЕМВС** - Embedded Microprocessor Benchmark Consortium – създава бенчмарк софтуер и хардуер за вградени системи. Продукти – AutoBench 1.1 (за процесори в автомобили, индустриални и общи приложения); ConsumerBench 1.1 (за цифрови фотоапарати, принтери и др.); DENBench 1.0 (процесори в мултимедийни системи); OABench 1.1 (процесори в принтери, плотери и др. офис системи) и др.
- **СПЕС** - Standard Performance Evaluation Corporation – за компютри. Продукти – SPEC CPU2006, комбинирана производителност на процесора, паметта и компилатора; програми за тестване на уеб сървъри, мейл сървъри и др.
- **ТПС** - Transaction Processing Performance Council – за измерване на обработката на транзакции и производителността на бази данни в компютърни системи за управление на стоки, резервации за авиолинии, електронно банкиране и др.

## **6. Популярни бенчмаркове**

- LavaLys Everest — тестване на подсистемите ALU и FPU на процесора, кеша, пропускателната способност на паметта, скорост на работа на твърдия диск
- FutureMark 3DMark — тест, използван за определяне производителността на компютъра в игровата тримерна графика
- PCMark — интегрален тест на всички подсистеми на компютъра
- 3D Robot FPS — тест, използван за определяне производителността на компютъра в тримерната графика и стабилността на системната памет.
- Aquamark — тест, използван за определяне производителността на компютъра в игровата тримерна графика
- ВАРСО SysMark — интегрален тест за производителността на компютъра в офисни приложения
- ВАРСО WebMark — интегрален тест за производителността на компютъра при създаване на Интернет съдържание
- ВАРСО MobileMark- тест за времето на автономна работа на портативни ПК, моделиращ различни видове натоварване
- Right Mark Memory Analyzer — тест за подсистемите на кеша и паметта.
- Right Mark Audio Analyzer — тестване качеството на аналоговите и цифрови канали на звукова апаратура
- CineBench — тестване на компютъра в режим на създаване на тримерна графика
- ScienceMark — тест на подсистемите ALU и FPU на процесора, кеша, пропускателната способност на паметта
- SiSoftware Sandra — съдържа модули за тестване на почти всички подсистеми на компютъра
- SuperPI — тест, изчисляващ числото Пи със зададено количество знакове след десетичната точка

- Business Winstone — интегрален тест за производителността на компютъра в офисни приложения
- Nero CD/DVD Speed — тестване на оптични устройства
- Lite-On IT K-Probe — тест за качеството на записваните оптични носители
- Exact Audio Copy — тестване на оптични устройства при работа с дискове CDDA (аудио CD)

### Скриншот от тестовата програма CineBench 2003



## 7. Бенчмаркове с отворен код

- **DEISA Benchmark Suite**: за определяне на производителността на паралелни суперкомпютърни системи (HPC High-performance computing) чрез изпълнение на реални приложения от широк кръг научни дисциплини: астрофизика, динамика на флуидите, моделиране на климата, биологични науки, материалознание, термоядрена енергия и физика на частиците.
- **Dhrystone**: производителност при целочислени аритметични операции
- **Fhourstones**: целочислен тест
- **HINT**: подрежда по производителност компютърните системи като цяло
- **Iometer**: измерва и характеризира входно-изходната подсистема за единични и клъстерни системи
- **Linpack / LAPACK**
- **Livermore loops**
- **NAS** паралелни бенчмаркове
- **PAL**: бенчмарк за физични двигатели в реално време
- **Phoronix Test Suite**: бенчмарк с отворен код за операционни системи Linux, OpenSolaris, and FreeBSD
- **POV-Ray**: 3D рендиране
- **TPoX**: бенчмарк за транзакции за XML бази данни
- **Ubench**: Прост бенчмарк за процесора и паметта за различни реализации на UNIX (вкл. Linux).
- **VMmark**: a server virtualization benchmark suite from VMware.
- **Whetstone**: производителност на системата при работа с дробни числа (с плаваща запетая).
- **LMBench**: прости преносими бенчмаркове, използвани за сравняване производителността на различни UNIX системи

## **8. Компютърните игри в ролята на бенчмарк**

Много компютърни игри, благодарение на тяхната взискателност към хардуера и зависимостта на скоростта на работа на играта от мощността на компютрите, успешно се използват като бенчмаркове. Такива игри са: Quake 2; Quake 3; Quake 4; Comanche 4; Return to Castle Wolfenstein; Serious Sam; Unreal Tournament 2003; Unreal Tournament 2004; Call of Duty; Thief: Deadly Shadows; Half-Life 2; Doom 3; Far Cry; Oblivion; F.E.A.R.; Condemned: Criminal Origins; Crysis; Unreal Tournament 3; Grand Theft Auto 4.

Обикновено тестването на компютъра в играта изглежда по следния начин: пуска се предварително записано тестово демо и се измерва количеството кадри в секунда (FPS), което е способна да предаде тестваната система. Резултатите от тестването с различни разделителни способности и с различни настройки на качеството се нанасят в таблица.

## **9. Ежедневните компютърни задачи в ролята на бенчмарк**

С голям успех за сравнение на производителността се използват различни задачи, ежедневно изпълнявани от потребителя, като:

- Използване на цифрови филтри в графичните редактори
- Компресия на файловете с използване на архиватори
- Кодиране на аудио- и видеофайлове
- Рендиране на сцени в 3D-редакторите

Измервайки времето за изпълнение на една и съща операция над еднакви изходни файлове, но на различни компютри, може да се получи сравнителна характеристика на производителността.

## **10. Бенчмаркинга като вид спорт.**

В последно време бенчмаркинга, т.е. състезаването по производителност на компютрите в специални тестове, се превърна в особен вид виртуален спорт. Основен тест, чрез който притежателите на мощни компютри се състезават помежду си е FutureMark 3DMark. На второ място по популярност у бенчерите е тестът Aquamark. Характерно за бенчмаркинга е разделянето на групи по интереси (притежаването на една или друга видеокарта, процесор и др.) и читърството (измамване на системата за регистрация на резултати).

Литература:

1. <http://www.benchmarkhq.ru>