Agenzia per l'Italia Digitale
Presidenza del Consiglio dei Ministri

Circolare AgID n. 06/2017

Circolare inerente i benchmark di sistema per piattaforme desktop, notebook e tablet basati su S.O. Microsoft Windows

Circular regarding system performance benchmarks for PCs Desktop, PCs Notebook, and PCs Tablet platforms based on O.S. Microsoft Windows

Raccomandazioni nell'utilizzo dei benchmark nelle procedure di gare pubbliche

Recommendations in using Benchmarks in the Public Procurement of Computers

Roma, 18 dicembre 2017

Preface

AgID (the Italian Digital Agency) has issued this Circular in order to provide the contracting administrations a reliable and intelligible tool to prepare those technical documents necessary for the official tenders to acquire of information and communications technologies products (ICT) in a non-proprietary manner, in accordance with the Legislative Decree dated 18th April 2016, nr. 50 (Code on public procurement) regarding the banning to mention brand, patent or product in the technical requirements of calls of tender.

In fact it should be possible to formulate current technical requirements of the ICT products to be procured without having to use specific product names, brand names or manufacturers.

This guarantees a non-discriminatory competition between suppliers which maintains the short-term and long-term efficiency of tenders and ensures their technical quality. **This guide covers desktop PCs, notebook PCs and Tablet-PCs**.

The national Act on public procurement, described above, forbids the use of brand names within the scope of public tenders, transposing anti-discrimination principles according to the EU procurement directive 2014/24/EU and EU contracts. The directives are aimed at preventing the type of discriminating wording in calls to tender that excludes particular manufacturers and suppliers from the group of potential bidders. As a result, whatever means a procurement authority chooses to describe system performance, it needs to ensure the tender specifications are non-discriminatory and don't exclude viable competitive vendors.

Glossary

Desktop PC: … is a personal computer designed for regular use at a single location on or near a desk or table due to its size and power requirements. The most common configuration has a case that houses the power supply (internal or external), motherboard (a printed circuit board with a microprocessor as the central processing unit (CPU), memory, bus, and other electronic components), disk storage (usually one or more hard disk drives and optical disc drives); and a keyboard and mouse for input. In this document, Desktop PCs include mini PCs and All-In-One PC. Thin clients are not part of this document.

Notebook PC: … is a small, portable personal computer with a "clamshell" form factor, an alphanumeric keyboard on the lower part of the "clamshell" and a thin LCD or LED computer screen on the upper portion, which is opened up to use the computer. In this document a notebook PC also includes other similar devices like detachable and 2 in 1 systems.

Tablet PC, or in short tablet, is a mobile computer with a touchscreen display, sensors (like screen orientation), digital cameras (back and/ or front including an accelerometer so images on screens are always displayed upright), microphone and speaker, processing unit, memory, storage, communication (Wi-Fi, GSM, Bluetooth) and a rechargeable battery in a single thin, flat package. The touchscreen display uses gestures by finger or stylus to replace the mouse, trackpad and keyboard. Please be aware that changes to the individual part are limited or not possible.

Stakeholders

AgID (the Italian Digital Agency) found necessary to substitute with this Circular those of CNIPA (CNIPA/CR/44 dated 5th October 2004 and CNIPA/CR/45 dated 27th December 2004 about "The Guidelines for the public supply of PC, notebooks and servers").

A pool of experts representing industry sector and benchmark developers has been consulted in the period starting from December 2016 until July 2017 about this goal. At the same time the support given by the Digital Transformation Team and Consip (The national Agency for public procurement) has guaranteed the correctness of the choices made respectively in terms of strategy for the digital growth and regarding the influence on reference markets.

The names of the different stakeholders and experts are listed in Annex nr. 2.

General Recommendations

Using benchmarks to assess performance

As computing technology has advanced, it has become more difficult to compare the performance of various computer systems simply by looking at their specifications. For example, a processor that operates at a higher clock frequency may not necessarily deliver more computational power. Since clock speed (frequency) does not necessarily predict performance when comparing processors with different architectures, from different manufacturers, tests (called benchmarks) have been developed to compare product performance. Relying only on physical parameters to compare different products may not result in a well-informed decision.

Benchmarks are specialized computer programs that run on the systems under evaluation. The benchmarking program executes a series of standard tests and trials simulating particular workloads on the system, and generates a final performance score. The performance score provides a snapshot of system performance on the workloads measured, which enables an objective, data-driven comparison.

Evaluating products using performance-based benchmarks, rather than processor numbers or clock speeds, can lead to better informed decisions. Benchmarks based on usage models can provide a solid framework for comparing performance of computing products to be deployed within government agencies.

Benchmark overview

There are numerous performance benchmarks available, and it is not always easy for a procurement authority to choose the most appropriate benchmark(s) for a specific tender.

Choosing an inappropriate benchmark could result in buying a computer system which is different from what the organization requires, and in some cases may even lead to discrimination against specific vendors and their products.

Regardless of which benchmark is chosen, it is extremely important to set up and follow a rigorous methodology when using performance benchmarks. *Variations in the way a benchmark is run may lead to the results being unreliable and not comparable,* which may even result in a challenge to the contract award.

The first aspect to consider is the type of benchmark. There are two types currently available in the market:

- **System benchmarks** evaluate the overall performance of a system for a defined usage model.

- **Component benchmarks** measure the performance of individual components, such as the CPU, memory, or graphics card.

For any of these types, a good performance benchmark should always have the *minimum attributes* described below:

A benchmark must:

- **Measure the overall performance** of the system and not just the performance of individual components so that it does not weigh the individual components disproportionately,
- Use **test scenarios** geared to the intended use,
- Represent **all relevant manufacturers and computer platforms**; its development process is independent and transparent,
- **Reflect**, in a **balanced manner, the performance** that can be expected during the term of use of the PC.

- **Recognised and built with stakeholder input**: Procurement authorities should choose performance benchmarks developed and maintained by well-recognized standards organizations with an independent, transparent and unbiased development process that takes all relevant industry stakeholder input into account.

- **Relevant and representative:** Government procurement authorities should choose a benchmark or a combination of benchmarks which measure performance using tests representative of the actual everyday use for which the system is intended.

  If the benchmark is not relevant, procurement authorities may risk purchasing a product which is different from what is needed.

- **Up-to-date:** Procurement authorities should always use the most recent version available of any given benchmark.

  Good performance benchmarks are continuously updated, and new benchmarks are regularly introduced to keep pace with development and innovation in the computer industry. A benchmark that is not up-to-date will not take into account how new features (e.g. multithreading) affect performance.

  Using an outdated benchmark to compare two systems may provide an inaccurate performance comparison. For example, assume a case where one system offers real performance benefits with improved technology, and a second, older system does not. If the outdated benchmark does not recognize and measure the technology improvement, the system with the old technology may actually score higher than the system with the new, better performing technology. In short, outdated benchmarks may disadvantage the most recent, innovative and better performing products.

Benchmark developers

Benchmark developers are categorised in order of transparency and openness as below:

- Non-profit benchmarking consortium (examples are BAPCo®, SPEC® and EEMBC®)

- Non-profit open source community (an example is Principled Technologies®)

- For-profit independent benchmark vendor (examples are Futuremark® and Kishonti® Informatics)

- Smaller for-profit developers (an example is AnTuTu®)

Website links:

BAPCo®   http://www.bapco.com/

SPEC®   http://www.spec.org/

EEMBC®   http://www.eembc.org/

FUTUREMARK® http://www.futuremark.com/

In the case of an open source community or independent benchmark vendor, the final decisions as to which features are included in a final released benchmark are typically made by the executives of the organization. Independent benchmark vendors vary in the level of user input they seek.

Benchmark recommendations

The workgroup recommendation is to look for a balanced system configuration which fits the needs on performance (great user experience), energy consumption and usage model.

Each benchmark listed in this Guideline is a complete and balanced tool whose use in public tenders is not to be assessed on the basis of the specific applications and libraries used inside the benchmark but only on its field of applicability.

The performance benchmarks recommended for desktop PCs, notebook PCs and Tablet PC are listed in Annex 1 followed by a detailed description.

Regarding the determination of benchmark scores to be set as a threshold for participation to tenders, AgID believes that this should be done based on the scores measured on systems available in the market at the moment of preparing the tender and meeting the needs of the procurement entity.

AgID will organize a workgroup with the most representative players of the market, at a minimum once every two years to update the Annex 1 based on technology evolution.

Digital signature
Antonio Samaritani

Allegato 1 – Raccomandazioni tecniche (giugno 2017)

Annex 1– Technical Recommendations (June 2017)

General Benchmark recommendations

The performance benchmarks recommended for desktop PCs and notebook PCs with Windows OS and Tablet PCs at the time of this document are listed below followed by a detailed description.

At the date these Recommendations don't cover Desktop PCs and Notebook PCs with MAC OS.

For desktop PCs:
- SYSmark* 2014 1.5 or later
- PCMark* 8 Work scenario (Conventional and Accelerated) or later
- 3DMark* (Graphic performance)

For notebook PCs:
- SYSmark* 2014 1.5 or later (Performance)
- MobileMark* 2014 1.5 or later (Battery Life)
- PCMark* 8 Work scenario or later (Conventional and Accelerated) (Performance)
- 3DMark* or later (Graphic performance)

For Tablet PCs:
- TabletMark* V3 / TabletMark*2017 (IOS/Android/Microsoft) or later

General Benchmark Descriptions:

SYSmark* 2014 1.5 is a benchmark from the BAPCo* consortium that measures the performance of Windows* platforms. SYSmark tests three usage scenarios: Office Productivity, Media Creation and Data/Financial Analysis. SYSmark contains real applications from Independent Software Vendors such as Microsoft* and Adobe*. Reported metrics: SYSmark 2014 Rating and a rating for each scenario result (higher is better for all). OS support: 32-bit & 64-bit Microsoft* Windows* 7, 8 and Windows* 10.

For a full list of applications, weighting and benchmarking methodology please refer to BAPCo's white paper for SYSMark 2014[1].

SYSmark* 2014 SE is the follow up of SYSmark* 2014 1.5. As addition to the three scenarios: Office Productivity, Media Creation and Data/Financial Analysis a new "Responsiveness" scenario is available. Workloads include: application launches, file launches, web browsing with multiple tabs, multi-tasking, file copying, photo manipulation, file encryption + compression, and background application installation also a new Energy Consumption feature which records energy usage in watt-hours during the workloads. OS support: 64-bit only Microsoft* Windows* 7, 8.1 and Windows 10.

PCMark* 8 is a benchmark from Futuremark* that measures Windows* everyday computing performance running on notebook-, desktop-, and tablet systems. The PCMark 8 Work benchmark test measures the system's ability to perform basic office work tasks, such as writing documents, browsing websites, creating spreadsheets and using video chat. PCMark*8 contain the real application LibreOffice Calc from the Document Foundation and the custom build workloads are using standard tools supported by Microsoft* and the Microsoft* Media Foundation. The Work benchmark is suitable for measuring the performance of typical office PC systems in scenarios which do not require media capabilities. PCMark 8 Work offers two running modes: Conventional and Accelerated. Reported metrics: Each

---

[1] https://bapco.com/wp-content/uploads/2015/09/SYSmark2014Whitepaper_1.0.pdf

test run reports a separate score (higher is better). Scores from different running modes (i.e. Conventional and Accelerated) are not comparable. In general, PCMark8 is offering six test scenarios: Home, Creative, Work, Storage, Application and Battery Life. Support for Microsoft* Windows* 7, 8.1 and Windows* 10.

•   In Conventional mode the workloads do not use OpenCL. This reflects how the majority of software works today, with the score providing a performance baseline.

•   "Run Accelerated" Accelerated mode allows workloads to use OpenCL acceleration, anticipating the way software will work in the future. For this test, the GPU (Graphics Processing Unit) is used in addition to the CPU (Central Processing Unit) for the video chat and the spreadsheet subtests.

For a full list of applications, weighting and benchmarking methodology please refer to Futuremark's Technical Guide for PCMark 8[2].

3DMark* is a benchmark from Futuremark* that measures DX* 9 / OpenGL* ES 2.0, DX 10 and DX 11 gaming performance running on notebook-, desktop-, and tablet systems. There are three main tests: "Ice Storm" for DX 9 / OpenGL ES 2.0, "Cloud Gate" for DX 10, "Sky Diver" for DX11 and "Fire Strike" for DX 11 graphics. Reported metrics: Graphics Score (GPU), Physics Score (CPU), Combined Score (GPU & CPU) and an overall 3DMark Score (higher is better for all Scores). Scaling efficiencies: Graphics tests are GPU dominant, sensitive to graphics and CPU frequency, core count and memory. OS support: Android*, iOS* and Microsoft* Windows*7, 8, 8.1 and Windows*10.

For a full list of applications, weighting and benchmarking methodology please refer to Futuremark's Technical Guide for 3DMark[3].

MobileMark* 2014 1.5 is a benchmark from the BAPCo* consortium that measures Windows* performance qualified battery life. MobileMark* provides two usage scenarios during battery rundown: Office Productivity and Media Creation. MobileMark* contains real applications from Independent Software Vendors such as Microsoft* and Adobe*. Reported metrics (per scenario): MobileMark 2014 1.5 Battery Life Rating and MobileMark* 2014 1.5 Performance Qualification Rating (higher is better for both). OS support: MobileMark* 2014 ver 1.5 supports 32 or 64-bit versions of Microsoft Windows 7, 8, 8.1 and Windows* 10.

For a full list of applications, weighting and benchmarking methodology please refer to BAPCo's white paper for MobileMark 2014[4].

---

[2] http://s3.amazonaws.com/download-aws.futuremark.com/PCMark_8_Technical_Guide.pdf
[3] http://s3.amazonaws.com/download-aws.futuremark.com/3DMark_Technical_Guide.pdf
[4] https://bapco.com/wp-content/uploads/2015/09/MobileMark_2014_White_Paper_v1.0.pdf

TabletMark* V3 is a benchmark from the BAPCo* consortium that measures User Experience for Windows* Modern UI light productivity, media & performance qualified battery life use-cases by incorporating common real-world tablet usages such as Web, Email, Photo, Video Sharing and Video Playback. The TabletMark V3 workloads exercise technologies such as HTML5/JavaScript*, AES 256 bit encryption, HDR imaging and H.264 video encoding. Reported metrics: TabletMark Rating, Web & Email Rating, Photo & Video Rating (higher is better for all), and performance qualified battery life. OS support: Android*, IOS* Microsoft* Windows* 8 and RT.

For a full list of applications, weighting and benchmarking methodology please refer to BAPCo's white paper for TabletMark* V3[5].

TabletMark* 2017 is the follow up of TabletMark* V3 with the same structure, new for 2017, it includes updated workloads, utilizes updated SDKs and development tools, and the Windows version has been rewritten from the ground up for Microsoft Windows 10 Universal Windows Platform.

For a full list of applications, weighting and benchmarking methodology please refer to BAPCo's white paper for TabletMark* 2017[6].

Performing SYSMark*, PCMark, MobileMark Benchmarks

For all benchmark values it must be assumed that fluctuations in the range of 2% – 5% will occur in each round of the test. Multiple measurements achieve greater accuracy. For this reason, at least three benchmark measurements should be taken. It may be necessary to adjust the number of test runs accordingly in the configuration menu of the benchmark. The arithmetic means used to calculate the test runs can be found in the benchmark protocol. Above and beyond the basic settings of the reference system, operating systems provide many setting options that may lead to significantly different results. We therefore recommend creating a protocol of the optional settings used and submitting the protocol along with the tender documents. In order to achieve comparable results, it is necessary to comply precisely with the specified procedure for performing the benchmark

Operating System Installation

For Correct Benchmarking
Do not install an »image«.
Take the following steps into consideration:
-    Enable and save the BOOT options (CD/DVD drive) in the BIOS.
-    HDD: format with NTFS on one partition.
     (This partition does not have to cover the whole disk.)
-    Install the OS from original source (DVD)
After the OS is installed:
-    Update all drivers (Chipset, LAN, USB, Graphic, Audio, USB, TPM……)
-    Use the device manager, checks must be conducted to ensure that all drivers are installed and all devices are listed (without error notifications): Start, Setting, Control Panel, System Properties, Hardware-Device Manager.

To conclude the installation, remove files:

---

[5] https://bapco.com/wp-content/uploads/2015/09/TabletMark-v3-WhitePaper-1.0.pdf
[6] https://bapco.com/wp-content/uploads/2017/02/TabletMark-2017-WhitePaper-1.0.pdf

- Start, All Programs, Accessories, System Tools, »Disk Cleanup« – enabling all items.
- If using magnetic hard drive space (HDD): Defragmenting of the hard disk:
- Run c:\defrag c: -f or Start, All Programs, Accessories, System Tools, Defragmentation.

Additional notes for all operating systems:

Before starting the benchmark run on the device, please check the following items: The required drivers have to be provided by the respective manufacturers and must be up to date.

NOT accepted optimizations are:

Overclocking of components through BIOS setup is not permitted. „
If possible, perform BIOS and firmware updates (via the Internet). The updates should be available from the Internet. „
Additional optimisation by using specific drivers or software and BIOS setups is not permitted.

After a successful installation of the OS install the selected benchmarking software and start the test (three runs are recommended.

Exception: For MobileMark battery life tests one run is accepted)

Configuration of SYSMark*

BAPCo* provides a configuration program which runs automatically at the start of each benchmark. No further settings are required.
Attention: please check (http://www.bapco.com) for patches.

The configuration settings will be selected and executed directly in the main program, BAPCo* Sysmark , under »CONFIGURE«.

The default standard values should not be changed.

Screen resolution: the screen resolution has only a minimal effect on the overall result, so this can be ignored. As a standard solution, with an aspect ratio of 4:3, a resolution of 1280x1024 is used.

Configuration of PCMark

Futuremark is not providing a configuration program.

Set the Windows power profile to »maximum performance«

Deactivate Windows update

Configuration of MobileMark for running the battery life benchmark:

BAPCo provides a configuration program which runs automatically at the start of each benchmark.
Attention: please check (http://www.bapco.com) for patches. „

The configuration settings will be selected and executed directly in the main program, BAPCo MobileMark, under »CONFIGURE«.

The default standard values should not be changed.

Wireless Connectivity
The Office Productivity and Media Creation scenarios require that the system's wireless network adapter be turned on and associated to a wireless network which is not connected to the Internet for the duration of the test.

Ambient Light Sensors
Use of adaptive brightness screen dimming technologies is strictly forbidden and must be disabled prior to running the test. Refer to Screen Brightness.

Screen Brightness
The test system display brightness must be set to a value not less than 150 nits, as measured on DC (battery) power in the center of an all-white screen. The setting must be maintained for the duration of the test, except when screen dimming, screen off or Connected Standby is in use (see below). In cases where the system display is not capable of reaching 150 nits brightness on DC (battery) power as measured in the center of an all-white screen, the screen brightness must be set to the maximum brightness value as measured on DC (battery) power, for the duration of the test, except when screen dimming is in use as per the rules below. System displays not capable of reaching 150 nits brightness on DC (battery) power must be documented

Screen Dimming
Use of operating system supported screen dimming capability is allowed at the
tester's option provided that:
☐ The installed Operating System is Microsoft Windows 7 or Windows 8.0*.
☐ The initial screen brightness is set as specified in the Screen Brightness section of this document.
☐ Screen dimming timeout value must be set to minimum of 2 minutes on DC power
☐ The dimmed display value must not be less than 45 nits, as measured on DC power
☐ *Screen dimming is not allowed in Windows 8.1 and Windows 10
Screen Off
Use of operating system supported screen off capability is allowed at the tester's option provided that:
☐ The installed Operating System is Microsoft Windows 8.1 or Windows 10*.
☐ Screen off timeout value must be set to minimum of 5 minutes on DC power.
☐ Screen off only occurs during idle periods of the benchmark, only after the point in each
   idle period where 5 minutes of idle has elapsed.
☐ *Screen off is not allowed in Windows 7 or Windows 8.0

Connected Standby
Use of operating system supported connected standby is allowed at the tester's discretion provided that:
☐ The installed operating system is Microsoft Windows 8.1 or Windows 10
☐ Screen off timeout value must be set to a minimum of 5 minutes on DC power
☐ Connected standby only occurs during idle periods of the benchmark, only after the point
   in each idle period where 5 minutes of idle has elapsed.
☐ The test system meets the requirements of the Microsoft Connected Standby specification. The
   requirements can be found here:

http://msdn.microsoft.com/en-us/library/windows/hardware/jj248729.aspx
- ☐ Standby/Hibernate: Use of operating system supported standby/hibernate capability is NOT allowed, with the exception of connected standby when used in accordance with the rules above.

Windows* 10 Battery Saver
Use of Windows 10 "Battery Saver" is permitted under the condition that it set to activate at no greater than 20% and the check box "Lower Screen brightness while in battery saver" is disabled Standby/Hibernate Use of operating system supported standby/hibernate capability is NOT allowed, with the exception of connected standby when used in accordance with the rules above

The performance part evaluation of the notebook PC evaluation can either be done with PCMark or SYSmark as per above.

Allegato 2: esperti e stakeholders consultati da AgID	***Annex 2****: list of experts and stakeholders has been consulted by AgID*

Alessandro Barbesta	ACER
Alessandro Casacchia	AgID
Antonio Cesarale	HP
Giulia Donati	HP
Jan Guetter	AMD
Nathan Harley	BAPCo
Silvia Invernici	INTEL
Ian Jones	INTEL
Andrea Luiselli	INTEL
Paolo Luxardo	CONSIP
Riccardo Magni	FUJITSU
Jukka Makinen	UL
Marcello Manca	UL
Roberto Mattioni	ASUS
Ulrich Norf	INTEL
Giampaolo Parravicini	AMD
Marco Pennetti	LENOVO
Marzia Pedrazzoli	DELL
Carmine Portanova	LENOVO
Ivan Renesto	DELL
Joerg Roskowetz	AMD
Gianluca Serrao	DELL
Dario Vallese	ACER