WHAT TO CONSIDER WHEN BENCHMARKING WORKSTATIONS

EXECUTIVE SUMMARY

Moor Insights & Strategy believes that benchmarks need to follow a specific set of rules to be useful. They must be up to date and relevant to specific workloads. The guiding principle of running any benchmark is that it must assess a system's performance in a particular application or workflow in a way that reflects a typical user experience. This means that a benchmark for professional workstation users should test professional applications and test them in a way that reflects day-to-day usage.

The benchmarking process should replicate entire workflows, not just one step in the workflow, with specific benchmarks testing different workflow segments using different methodologies. The best benchmarks take actual applications and automate the workflow using the same files to generate realistic, reproducible results without any surprises.

Although benchmarking tools for consumers and "prosumers" (freelancers and individual workers) are often quick and easy to use, they are not sufficient for professional markets because of their very narrow testing capabilities.

There is still a place for synthetic benchmarks for quick comparisons; however, buyers should not make decisions based on one benchmark alone. Sometimes that means combining multiple benchmarks and comparing multiple scores to make an assessment. We believe benchmark suites take the most appropriate approach. This research paper will explore the different benchmarking tools available across the workstation PC industry and evaluate their efficacy in informing workstation buyers' decision processes.

WORKSTATION MARKET CONDITIONS

Several trends have driven the workstation market toward continued growth, including significant performance improvements thanks to the adoption of GPU-accelerated rendering and the increased use of real-time graphics. Although COVID-19 impacted workstation growth in 2020, the industry saw significant growth in 2019, setting record

shipment numbers of over 1.7 million units in Q3 2019 and 18%+ growth year over year[1].

Workstations themselves are a sub-segment of the PC market and, for the sake of this paper, we define them as PCs used for the express purpose of completing a compute-heavy task for an employer or a customer. Workstations usually fall in a few major industry verticals: AEC (Architecture Engineering and Construction), Scientific/Research, Media & Entertainment, Surveillance, and Oil & Gas.

The workstation market is relatively mature, and the market players are all relatively well-known. Some of the most notable workstation software companies include Adobe, Autodesk, Avid, Dassault Systemes, Esri, Maxon, PTC, and Siemens. These companies often have multiple professional applications that knowledge workers use, which adds to the complexity of qualifying a workstation's stability and performance. They also regularly release new versions of their applications that need to be qualified on existing workstations to ensure they do not affect performance and stability. This qualification process generally results in longer lead times between software updates.

In addition to the software vendors, you have the hardware vendors that work with the software companies to validate the combination of the hardware and software and certify those test results. Roughly 90%[2] of the workstation hardware market consists of three companies: Dell, HP, and Lenovo. Other players like Boxx, Puget, Eurocom, Maingear, Silverdraft, and Velocity Micro are boutique builders that offer more custom solutions. In addition to these workstation OEMs, a few component suppliers drive the hardware platforms, as much of the processing happens on the CPU and GPU. Intel and AMD round out the CPU offerings while NVIDIA and AMD round out the GPU offerings, making for a reasonably small yet influential group of companies.

In general, the workstation market is still relatively small compared to the rest of the PC market. But its higher margins and average selling prices (ASPs) make it an attractive market for companies willing to invest the resources to pass the certifications necessary to compete. Because this market is highly lucrative for those who can compete in it, there is an added incentive to claim performance leadership among workstation OEMs and their component suppliers.

Since the system CPUs and GPUs are where most of the application performance comes, they are generally the focal points of most performance metrics. Many industry

---

[1] 2019 Was Quite a year, 2020 Will Be Even Greater
[2] Tier 1 Workstation Vendors vs. Boutique Suppliers: Different Strokes for Different CAD Folks - Cadalyst

benchmarks exist to measure performance; however, there is confusion about what makes a good benchmark and how it benefits the professional user experience.

RESEARCH

Benchmarks are fundamental tools that the industry uses to back up performance claims by component vendors, software vendors, and workstation OEMs. Company use of benchmarks varies in the workstation market, especially with the different types of benchmarks. We classify benchmarks into three categories: internal benchmarks and two types of industry-standard benchmarks. These benchmarks serve different purposes and have varying levels of validity depending on the audience and application workload.

INTERNAL BENCHMARKS

A company that uses professional workstations and related software in a particular way typically develops internal benchmarks. The company's CTO or technical staff will often create parameters and workflows that match employees' work. Not all companies are the same, of course, and application workloads and workflows vary, as do priorities. The CTO ultimately decides how to set up, test, and keep the internal benchmarks up to date to match the latest software versions. In many cases, however, this requires a significant number of resources from internal staff. These internal benchmarks generally lag behind the latest available versions of software and change with hardware and component architecture changes.

INDUSTRY-STANDARD BENCHMARKS

Industry-standard benchmarks generally come in two forms, consumer and professional. Some of the consumer benchmarks do have professional versions, but they are generally for OEMs to validate performance so that when a consumer uses the benchmark, the system performs as expected. For this paper, we will refer to consumer benchmarks as consumer/prosumer as many of these benchmarks cater to a student and hobbyist crowd but can at times appeal to freelancers and individuals who work as single employees.

CONSUMER/PROSUMER BENCHMARKS

We will quickly cover some of the more popular and useful consumer/prosumer benchmarking tools. Although they can be quick and easy to use, they usually only test specific workloads, and industry professionals do not generally use them.

TABLE 1: CONSUMER/PROSUMER BENCHMARK RESEARCH

Consumer/Prosumer Benchmarks	Gaming	Digital Content Creation (M&E)	Component Test	Office Productivity	Test Suite?
Cinebench		✓			✗
HandBrake		✓			✗
PerformanceTEST			✓		✓
Octane Bench		✓			✗
V-Ray 5		✓			✗
Blender		✓			✗
SiSoft Sandra			✓		✓
AIDA64			✓		✓
PCMark 10		✓		✓	✓
3DMark	✓				✓

(Source: Moor Insights & Strategy)

- **Cinebench**[3], a very popular prosumer benchmark, is a derivative benchmark of Cinema4D, a popular 3D rendering application for visual effects (VFX) artists made by Maxon.
- **HandBrake**[4] is an application and benchmark that consumers and freelancers very commonly use to evaluate the video transcoding capabilities of a PC based on the hardware encoders available from Intel, AMD, or NVIDIA.
- PassMark **PerformanceTEST**[5] is a suite of performance tests that are very user-friendly and designed to get relative performance numbers based on CPU tests, 2D graphics tests, 3D graphics tests, system disk tests, and memory tests.
- OTOY's **OctaneBench**[6] and Chaosgroup's **V-Ray 5** [7]are similar benchmarks that test a particular 3D workload, allowing the user to choose between CPU or GPU benchmarking. However, they do not necessarily test a specific workflow, even though they test a real-world workload in this sense.
- **Blender**[8] is an open-source 3D creation application gaining momentum in the industry. Blender has developed its benchmark with functionality similar to the OctaneBench and V-Ray 5 benchmarks.

---

[3] Cinebench R23 - Maxon
[4] HandBrake – The open source video transcoder
[5] PerformanceTEST – Easy PC Benchmarking for Windows – PassMark Software
[6] OctaneBench 2020.1 - OTOY
[7] V-Ray 5 Benchmark - Chaosgroup
[8] Blender Open Data – Blender Benchmark

What to Consider When Benchmarking Workstations   April 2021
Copyright ©2021 Moor Insights & Strategy

- **SiSoft Sandra**[9] and **AIDA64**[10] are component-level benchmarks that test particular CPU and GPU workloads and memory and storage. However, these benchmarks aggregate the performance numbers of different components, giving them a weighted score. In that sense, these benchmarks are similar to PassMark's PerformanceTEST.

- UL Benchmarks' **PCMark 10**[11] is a complete system benchmark that allows the user to test performance, battery, or storage. It makes available shortened and extended versions of the benchmark. PCMark 10 uses open-source software to simulate real-world application usage, including productivity software, video streaming, digital content creation, and web browser performance.

- UL Benchmarks also makes a very popular cross-platform gaming benchmark called **3DMark**[12], which has been around almost as long as the company. Most people use 3DMark to assess and compare performance to other graphics cards, either integrated or discrete, although it is a gaming benchmark first and foremost.

PROFESSIONAL BENCHMARKS

Professional benchmarks generally use multiple workloads within professional applications and are designed to replicate actual user workflows. A consortium of experts who understand the workstation market, applications, and users designed many of these benchmarks.

Cadalyst Benchmark[13] is a benchmark developed by the *Cadalyst* magazine, which has been around for 35 years and understands the workstation market very well. Cadalyst Benchmark only benchmarks one application, AutoCAD, which is essential to the engineering and education communities within the workstation market. However, the benchmark is currently on version 2015, utilizing AutoCad 2016, and does not appear to have been updated since 2018. Autodesk released AutoCAD 2021 in March 2020, so Cadalyst, while free, is dated.

---

[9] SANDRA 20/21 - SiSoftware
[10] AIDA64 - Finalwire
[11] PCMark 10 – The Complete Benchmark
[12] 3DMark – 3DMark.com
[13] Cadalyst Benchmark Test – Test Your Hardware Systems Running AutoCAD 2000-2019

TABLE 2: PROFESSIONAL BENCHMARK RESEARCH

Professional Benchmarks	AEC	Digital Content Creation (M&E)	Economy/ Finance	Scientific	Data Science	Other	Test Suite?
Cadalyst	✓						⊗
SPEC WorkStation3		✓	✓	✓	✓	✓	✓
SPECviewperf	✓						✓
SPEC SolidWorks 2020	✓						✓
Procyon		✓					✓
MLCommons					✓		✓
PugetBench		✓					⊗
PCMark 10 (Apps)						✓	✓

(Source: Moor Insights & Strategy)

Having provided objective benchmark results for computing for more than 30 years, SPEC (Standard Performance Evaluation Corporation) is one of the industry's most well-known and respected professional workstation benchmarking companies. The company builds a suite of benchmarks for different applications and publishes its tested workloads and methodologies.

SPECworkstation 3[14] is the latest version of the workstation benchmark, focusing primarily on CPU and system performance. It uses a broad array of professional and prosumer applications in each of the different workstation verticals mentioned earlier, ensuring that each vertical gets its own set of benchmarks and workloads.

In addition to SPECworkstation, SPEC also makes another workstation-relevant benchmark called **SPECviewperf**[15], which is currently in its 14th iteration. SPECviewperf takes a subset of the applications that are in SPECworkstation and explicitly tests the workstation's graphics processing capabilities in professional applications like 3ds Max, CATIA, Creo, Maya, Showcase, Siemens NX, and Solidworks.

Additionally, SPEC has created energy and medical benchmarks using open-source software to replicate workloads specific to those vertical markets and makes an application-specific benchmark for **SolidWorks 2020,** but that is only one benchmark for one workstation application.

---

[14] SPECworkstation 3 benchmark
[15] SPECviewperf 2020

What to Consider When Benchmarking Workstations   April 2021
Copyright ©2021 Moor Insights & Strategy

SPEC also has CPU-focused benchmarks, namely SPEC CPU which includes the SPECint and SPECfp benchmark suites. Although they do include multiple CPU workloads, they are too narrow and do not replicate real-world workflows or entirely workstation-focused workloads. SPEC's 2006 version of this benchmark is popular in the industry but outdated and does not replicate actual workstation workloads. SPEC CPU 2017 is not as commonly used but does include a newer version of these benchmarks with architectural improvements.

Along with the mostly consumer-focused 3DMark and prosumer benchmark PCMark10, UL Benchmarks also makes a new workstation benchmark called **Procyon**[16], which uses Adobe's suite of creative applications (Photoshop, Lightroom, and Premiere Pro) to benchmark creative performance for photo and video editing. Procyon requires an Adobe license, but this also ensures that the benchmark is running on the latest versions of these applications and is testing the actual application performance rather than a proxy.

Although not a nonprofit consortium like SPEC, UL Benchmarks still engages and works closely with the entire industry to ensure that all vendors have an even footing to compete in the benchmarks. UL Benchmarks also makes an AI benchmark as part of the Procyon family but focuses on Android smartphones rather than PCs.

MLCommons,[17] formerly MLPerf, is a newcomer to the market. It aims to give users an idea of what kind of machine learning or AI performance a particular processor or system can attain. Currently, MLCommons supports both machine learning training and inference benchmarks, with a heavier focus on inference with workloads that range from low-power processors to datacenter. As far as workstation users go, training and inference are relevant for different vertical workstation markets. As machine learning continues to grow, however, this benchmark may not be appropriate for data scientist workloads due to the differing types of workloads.

Other benchmarks do exist, like **PugetBench,**[18] primarily for the media and entertainment market, from system OEM Puget Systems. This benchmark includes Adobe's suite of creative software like Photoshop, Lightroom, Premiere Pro, After Effects, and AErender. PugetBench also includes DaVinci Resolve, which originated as a color grading application and has evolved into a video editing suite like Premiere Pro.

---

[16] UL Procyon – Professional Benchmark Suite
[17] MLCommons – Machine Learning innovation to benefit everyone
[18] PugetBench – Real-world Benchmarks to improve your workflow

PugetBench is the closest to what we would consider an in-house/internal professional benchmark suite, which is something to consider when testing a workstation's performance. In-house benchmarks like Puget's require more time to select applications and software versions relevant to workstation users. In-house or internal benchmarks need to consider best practices and how to replicate workloads, workflows, and reliability of results. This usually means creating test workloads that exist purely for benchmarking purposes and running them at least three times in a row to account for thermal variance and average results.

BENCHMARK ANALYSIS

The PC hardware industry leans toward more straightforward and easier-to-use benchmarks like Cinebench and Blender, which do not accurately represent a workstation's performance. Yes, it tells us how well the system will run a render in Cinema4D and Blender, but it does not show how the application loads or overall workflow performance.

Applications like V-Ray Bench and OctaneBench are mostly plug-in-based, so they generally focus on the rendering phase of work rather than the complete workflow. HandBrake similarly suffers from this as it only accounts for the transcoding component of a video editing workflow. It is true that rendering 3D images and video can sometimes be one of the most time-consuming components in a workflow, but it isn't the complete workflow.

Component benchmarks like PerformanceTEST, AIDA64, and Sandra aim to give system performance results but have somewhat opaque scoring systems and don't replicate professional workloads or workflows. These can prove useful for comparing individual components outside of the workstation industry but should not be used for measuring workstation performance.

While the Cadalyst benchmark is quite good in its focus, it is also out of date and lacks relevance. A benchmark like PCMark 10 may replicate real-world workflows and applications, but unless you run the PCMark 10 Applications benchmark, it is probably out of date, having come out in 2017. Additionally, PCMark 10's workloads are not necessarily focused on professional workstation users as they aim toward office worker productivity. PCMark 10 Applications does run Microsoft Office apps and workflows, but isn't commonly used by most people when running the benchmark and it isn't included as the primary benchmark suite.

What to Consider When Benchmarking Workstations

Consumer/Prosumer Benchmarks	Gaming	Digital Content Creation (M&E)	Component Test	Office Productivity	Test Suite?
Cinebench		✓			✗
HandBrake		✓			✗
PerformanceTEST			✓		✓
Octane Bench		✓			✗
V-Ray 5		✓			✗
Blender		✓			✗
SiSoft Sandra			✓		✓
AIDA64			✓		✓
PCMark 10		✓		✓	✓
3DMark	✓				✓

Professional Benchmarks	AEC	Digital Content Creation (M&E)	Economy/ Finance	Scientific	Data Science	Other	Test Suite?
Cadalyst	✓						✗
SPEC WorkStation3	✓	✓	✓	✓	✓	✓	✓
SPECviewperf	✓						✓
SPEC SolidWorks 2020	✓						✓
Procyon		✓					✓
MLCommons					✓		✓
PugetBench		✓					✗
PCMark 10 (Apps)						✓	✓

(Source: Moor Insights & Strategy)

Machine learning is still relatively new, and those workflows and applications are still developing and refining. MLCommons, while an AI and ML benchmark, isn't necessarily designed for workstation applications. There are not many industry-accepted and trusted machine learning benchmarks for workstations used by data scientists; most of the benchmarks out there today are available on GitHub and test individual components of a data scientist's workflow like Python or TensorFlow performance.

Good benchmarks reflect more professional workflows and compute-intensive workloads, as we covered earlier. Therefore, we don't necessarily see benchmarks like Cinebench as good workstation benchmarks, even if Cinema4D is the only application you use professionally. Workflow benchmarking is vital because it measures the time it takes to render a scene or an object, load the application, load assets, and make

changes before a final render begins. The best benchmarks take actual applications and automate the workflow using the same files, which generates realistic, reproducible results without any surprises.

Synthetic benchmarks do have their place in the industry, namely for making quick comparisons of specific workloads. Basing an entire buying decision on one of these benchmark tests, however, is an opportunity for error. Benchmark suites are more appropriate benchmarks as they combine different workloads and attempt to replicate a workflow. These suites can combine different components of a workflow through different workload tests from different benchmarks that would otherwise not present the complete picture on their own.

Based on our benchmarking principles, we believe that SPECworkstation 3 and UL Benchmarks' Procyon Content Creation benchmarks are the best evaluators of workstation performance.

SPECworkstation 3 provides value because it includes workloads from many workstation industry verticals and tests many different industry-relevant applications. SPEC's SPECviewPerf 2020 is a graphically focused benchmark and covers far fewer applications, so we are less inclined to recommend it unless assessing only GPU performance is essential. SPECworkstation 3 tests many of the same applications.

UL Benchmarks Procyon Content Creation is also a benchmark that we can recommend because it leverages Adobe's creative suite of applications to run and replicates the actual user workflow rather than just testing one part's workload.

CONCLUSION

Workstation users have a very particular set of needs that include compatibility with professional applications and rock-solid stability. There are many benchmarks out there that do not accurately address those needs. A good benchmark should be up to date and relevant and account for users' workflows. Few benchmarks out there test the complete workflow for users, and as a result, the users get short-changed when they purchase a system expecting a certain level of performance.

Ultimately, companies should test applications as they see fit but should know that many benchmarks only test one component of a workflow and should not be used in isolation. Companies should take a holistic approach when benchmarking to account for the many different types of users and applications. Building and maintaining your

benchmarks isn't easy or cheap, and it makes sense to use the few industry benchmarks that are up to date and well-maintained to make wise decisions.

IMPORTANT INFORMATION ABOUT THIS PAPER

CONTRIBUTOR
Anshel Sag, Technical Writer at Moor Insights & Strategy

PUBLISHER
Patrick Moorhead, Founder, President, & Principal Analyst at Moor Insights & Strategy

INQUIRIES
Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

CITATIONS
This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

LICENSING
This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

DISCLOSURES
This paper was commissioned by Intel. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events. Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.