SERVERMARK™

Technical Guide

Updated April 19, 2018

This guide updated April 19, 2018

SERVERMARK

Servermark is a UL benchmark suite developed by the same team that makes *3DMark*, *PCMark*, and *VRMark* benchmarks and *Testdriver*, a benchmark automation solution.

Designed for industry, enterprise and press use, Servermark brings UL's trusted expertise to IT professionals searching for accurate and impartial server benchmark tests.

Servermark tests are easy to install, run, and understand, which means you can choose and compare server hardware configurations with confidence.

SERVERMARK BENCHMARKS AT A GLANCE

The current release of Servermark includes the Media Transcode benchmark. More Servermark benchmarks are in development and will be available with future updates. Each of these benchmarks is designed for a specific type of server use.

BENCHMARK	AVAILABILITY
Media Transcode	Current release
Virtual Desktop Infrastructure	Coming soon
Web Server	In development
File Server	Planned
Database Server	Planned

LATEST VERSION NUMBERS

	VERSION
Servermark Controller	1.0.433
Servermark Media Transcode benchmark	1.0

Controller version number

The Controller version number applies to the software as a whole. It may change often as we update the application to add new features and ensure compatibility with the latest hardware. We recommend using the latest Controller version.

Benchmark version numbers

A benchmark version number is specific to a test. For example, the Media Transcode benchmark has its own specific version number. Benchmark version numbers change rarely and only when absolutely necessary to accommodate changes in third-party applications or bug fixes.

Comparing scores across versions

UL guarantees that benchmark results are comparable across Servermark Controller versions provided that the major digit of the benchmark version number is the same.

SERVERMARK OVERVIEW

Servermark consists of two components: the Controller and the Client.

The Controller and the Client work together to run the desired benchmarks on the system under test.
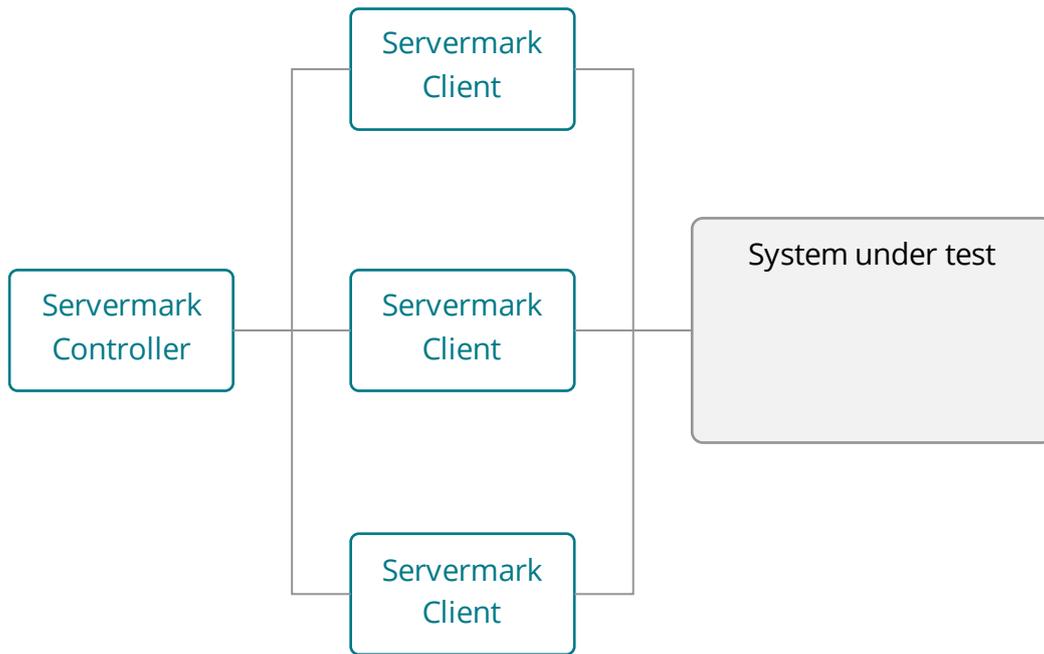
Servermark Controller

The Servermark Controller is a Windows application that can be invoked from the graphical user interface or from the command line. During a benchmark run, the Servermark Controller communicates with Servermark Clients to perform the desired tests on the system under test.

Servermark Client

You can run the Servermark Client as a Windows command line application or as a Linux console program. It can be installed on the system under test or on a separate *helper computer,* which is then used to generate a load heavy enough to reach the server's limits. In some cases, several *helper computers* may be needed as shown in the diagram on the next page.
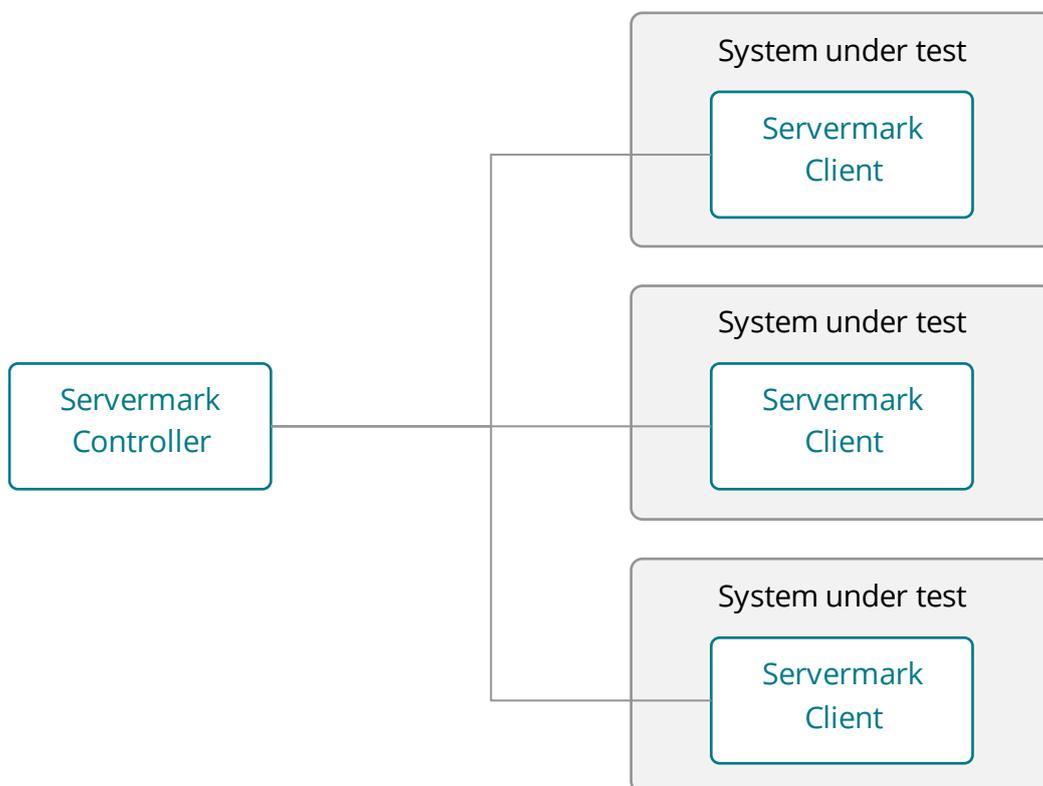

```
Installing the Servermark Client on the system under test
```

Installing the Servermark Client on separate PCs

SERVERMARK CONTROLLER

Minimum requirements

OS	Windows 7 Service Pack 1, 32-bit or 64-bit
CPU	Dual core
SYSTEM MEMORY	2 GB
STORAGE	10 GB free space
NETWORK	A network connection to Servermark Clients

Installing the Servermark Controller

To install the Controller, run `servermark-setup.exe`, which you can find inside `Servermark-{version}.zip`. Please make sure that the computer meets the minimum requirements above before installing, and note that Administrator rights are required. Once installed, open the Servermark application and enter your license key to unlock your benchmarks.
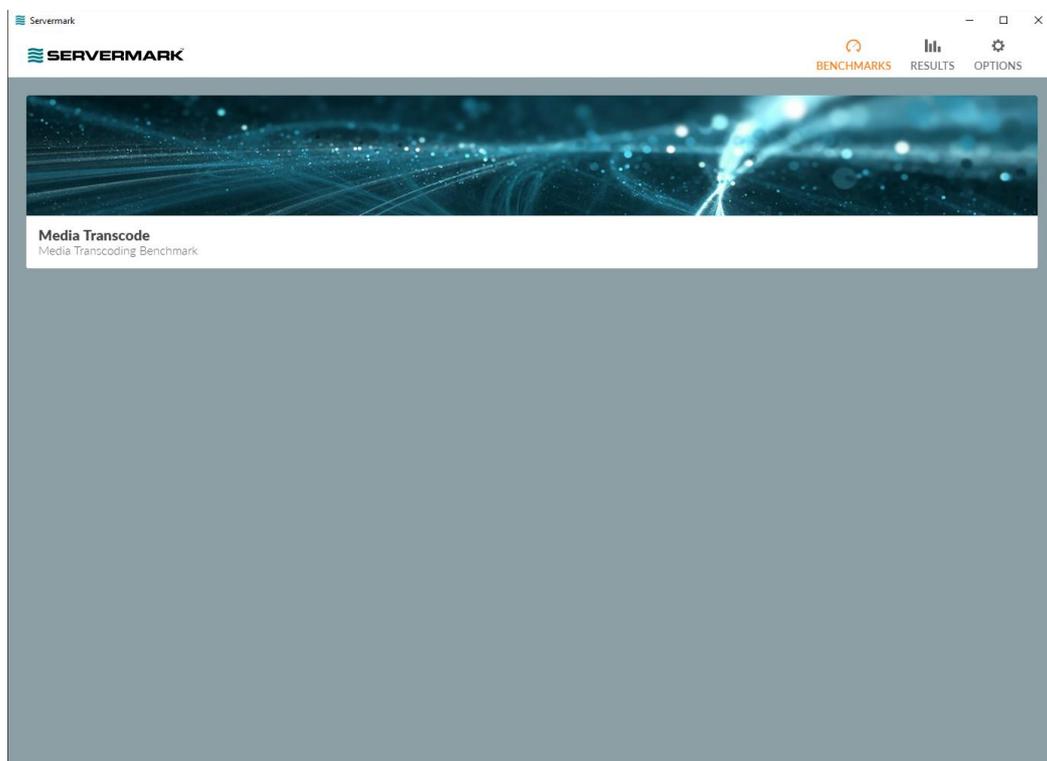
Adding Servermark Clients

Before you start benchmarking, you need to add Clients to the Controller. On the **BENCHMARKS** screen, click on a benchmark that is available with your license. On the next screen, you will see the options to search for Clients or add them manually.

Client search

The **First IP Address** and **Last IP Address** fields set the range for the search. The **Port** field is the TCP port used for communication between the Controller and the Client. The default port is 10003. Press **RUN SCAN** when ready and any Clients found within the range will be added.

Insert a Client manually

To add a Client manually, enter its location into the **IPv4, IPv6 or Hostname** field. The **Port** field is the TCP port used for communication between the Controller and the Client. The default port is 10003.

Press **ADD CLIENT** when you are ready and the Client will be added. You will see an error message if the Client could not be found.

Options screen

You can see and manage your license information on the **OPTIONS** screen.
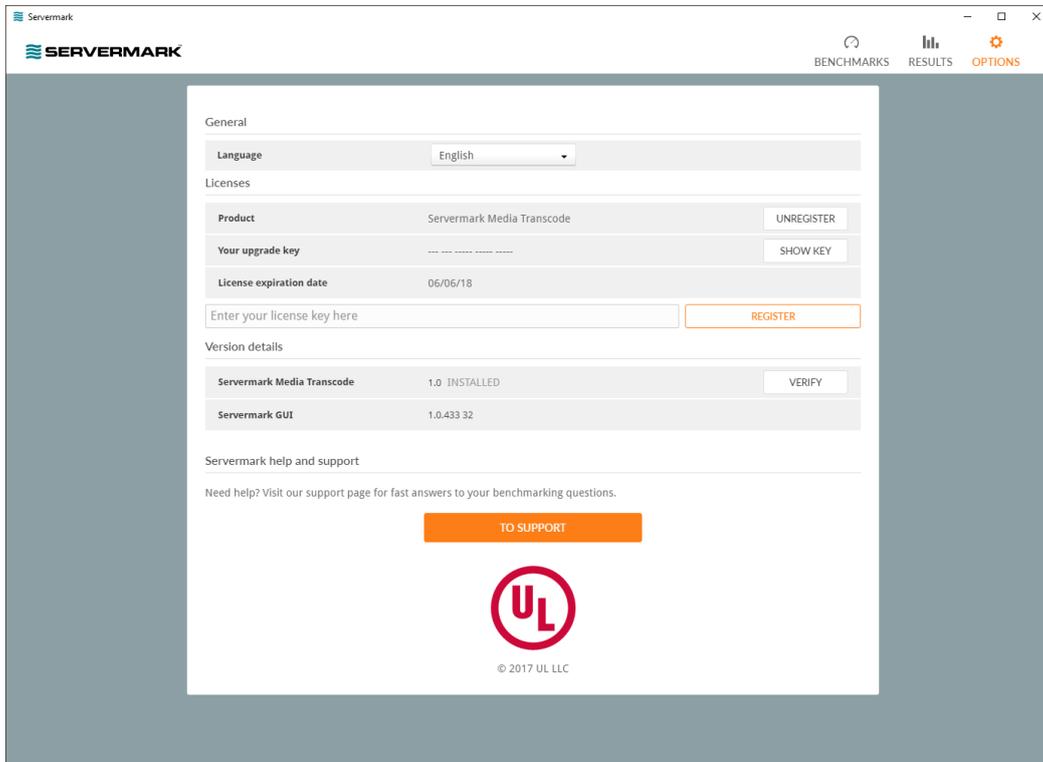
General

Choose a language from English, German, Russian, and Simplified Chinese.

Licenses

This section shows which benchmark licenses are activated and when they will expire. If you wish to unregister your key, so you can move your license to a different computer for example, press the **UNREGISTER** button.

Version details

Here you see the current version number of the Servermark Controller application and your benchmarks. If a newer version is available, you will be able to update from this screen.

Servermark help and support

If you need it, click on the **TO SUPPORT** button to contact our support team.

MEDIA TRANSCODE BENCHMARK

Media Transcode is a Servermark benchmark test for media transcoding servers. The benchmark can be used to test the performance of media transcoding servers individually or as a cluster.

The benchmark aims to determine how many video streams the server(s) can transcode simultaneously and provide you with tools to compare video quality. The benchmark uses high-performance media transcoding libraries from Intel and NVIDIA to perform the transcoding.
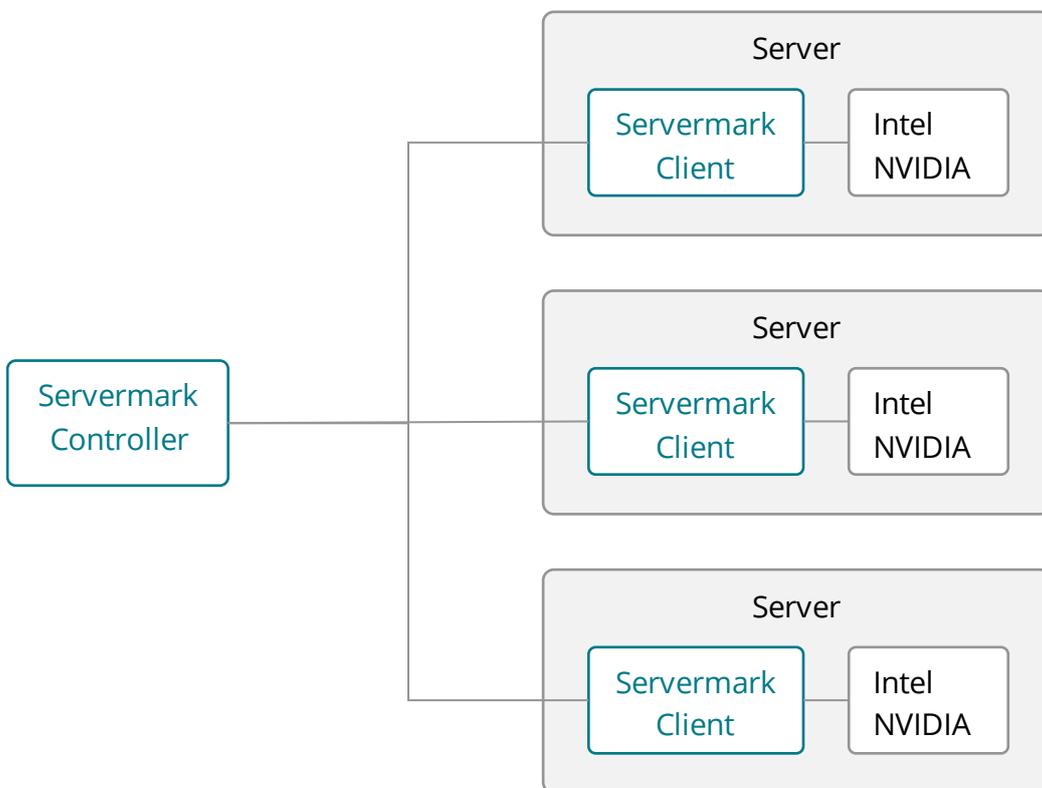
SERVER REQUIREMENTS

Intel workload

OS[1]	CentOS Linux 7.2.1511
CPU[2]	Intel Xeon E3-1200 v4 family Intel Xeon E3-1200 v5 family with C236 chipset Intel Xeon E3-1500 v5 family with C236 chipset 5th Gen Intel Core processor required for AVC 6th Gen Intel Core processor req. for AVC & HEVC
SYSTEM MEMORY	8 GB
STORAGE	100 GB free disk space
DRIVERS	Intel Media Server Studio 2017 R2
DEPENDENCIES	openSSL

---

[1] CentOS Linux 7.2.1611 will be supported in the next Servermark release.

[2] Note that Intel Kaby Lake processors are not supported by Intel Media Server Studio SDK 2017 R2.
See the release notes for details.

NVDIA workload

OS	Ubuntu Linux 16.04 LTS Later releases are likely to work but are untested
CPU	Intel Xeon, Core i5/i7/i9 or AMD Ryzen CPU with at least 4 physical cores
GPU	A second-generation Maxwell (GM20x), a Pascal-architecture NVIDIA graphics processor, or later
SYSTEM MEMORY	8 GB
STORAGE	100 GB free space
DRIVERS	NVIDIA GeForce 375.20 graphics card drivers or newer
DEPENDENCIES	openSSL

PREPARING THE SERVER

Install Intel drivers

If you are planning to run the Intel workload you must install the Intel Media Server Studio 2017 R2 runtime. The community version is available from [Intel's website](#).

On CentOS 7.2.1511, follows these steps with `Servermark-{version}.zip` in the `MediaTranscodeBenchmark` folder:

```
tar -zxf ServermarkMediaTranscodeBenchmarkClient-*.tgz
cd ServermarkMediaTranscodeBenchmarkClient
cd centos
cd install_community_intel_sdk
sudo ./install_community_intel_sdk.sh
reboot
```

Install NVIDIA drivers

If you are planning to run the NVIDIA workload you need only install NVIDIA graphics card drivers. On Ubuntu 16.04, follow these steps:

```
sudo add-apt-repository ppa:graphics-drivers/ppa
sudo apt-get update
sudo apt install nvidia-375
```

Install dependencies

The Servermark Client uses OpenSSL to calculate the checksum. If OpenSSL is missing, you can install it using package manager.

Servermark uses FFmpeg to calculate the peak-signal-to-noise ratio (PSNR) between the original and the transcoded video. PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation.

Please make sure you have FFmpeg binary in the search path by running the command `FFmpeg -version`.

On CentOS, you can download a pre-built static FFmpeg binary from

```
https://www.johnvansickle.com/FFmpeg/
```

Extract the downloaded package to your preferred location.

```
tar -Jxf FFmpeg-release-64bit-static.tar.xz
```

If needed, add it to your PATH environment variable by adding the following line to file `~/.bashrc` of the user who will be running the workload:

```
export PATH=$PATH:/path/to/FFmpeg
```

On Ubuntu 16.04, you can install FFmpeg with package manager:

```
sudo apt install FFmpeg
```

On both CentOS and Ubuntu, you can also pull down the FFmpeg source code from [FFmpeg code repository and build it locally](), but using the package manager or a pre-compiled static binary is generally an easier and faster way to get started.

Firewall exception

Verify that the port the Servermark Client is going to use is open.

On CentOS 7.2, the port can be opened as follows:

```
firewall-cmd --permanent --add-port=10003/tcp
firewall-cmd --reload
```

On Ubuntu 16.04, use:

```
ufw allow 10003
```

INSTALLING THE SERVERMARK CLIENT

Before installing the Servermark Client, make sure that the transcode server meets the [requirements](#) for running the workload.

The Servermark Client executable can be found inside the `ServermarkMediaTranscodeBenchmarkClient-{version}.tgz` package in the `MediaTranscodeBenchmark` folder in the `Servermark-{version}.zip`.

Copy the file to each Client machine and run this command:

```
tar -zxf ServermarkMediaTranscodeBenchmarkClient-{version}.tgz
```

Then, navigate to the installed operating system folder and start the Servermark Client using one of the following commands as appropriate:

```
./mt_intel
```

or

```
./mt_nv
```

After the Client has started successfully, you should see output like this:

```
[user@localhost media-transcoding]$ ./mt_intel
Command line: ./mt_intel
Servermark - Media transcode
Workload version: x.y.z.k
Waiting for a new client on TCP port: 10003
```

> ⚠ If the Intel workload does not run, check the CentOS graphic driver version. If it is *Gallium*, please verify that your BIOS secure boot is set to the *Other OS* position, otherwise CentOS may not be able to load the drivers.

The Servermark Client accepts following start parameters.

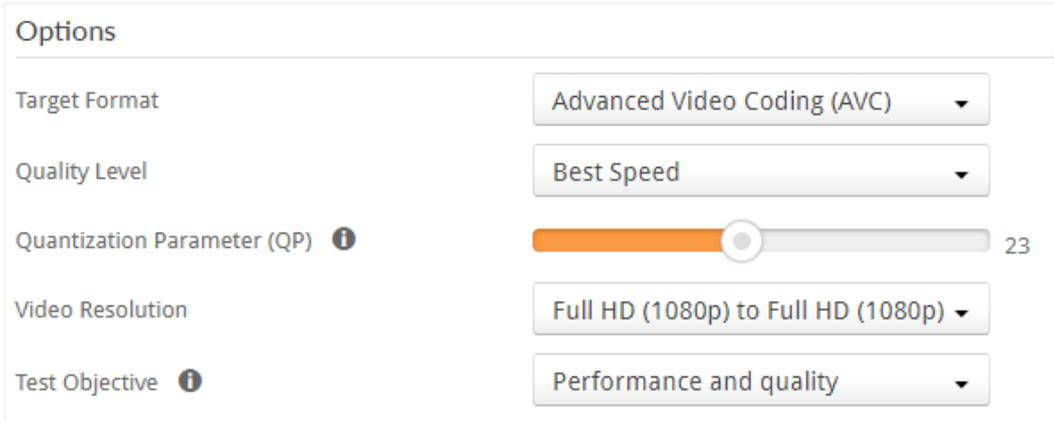| ARGUMENT	DESCRIPTION
-P PORT	The TCP/IP port on which the Servermark Client accepts connections (default: 10003)

RUNNING THE BENCHMARK

Before running the benchmark, make sure that the [Servermark Controller is installed](#) and that the [Clients have been added](#).

Options

Open the Servermark Controller application. On the **BENCHMARKS** screen, click on the Media Transcode box. On **MEDIA TRANSCODE** screen, use the **Options** section to choose and configure the benchmark settings.

Target Format

The Target Format setting specifies the video format for the output videos. The default setting is **Advanced Video Coding (AVC)**.

- Advanced Video Coding (AVC)
- High Efficiency Video Coding (HEVC)

Quality Level

The Quality Level setting specifies the output quality that the encoder will use. The default setting is **Best Speed**.

- Best speed
- Best quality

Quantization parameter

Quantization parameter (QP) has a significant effect on the compression rate. The range of the parameter setting is 0-51. The default value is **23**.

Lower values result in better quality at the expense of larger file sizes. Higher QP values increase the amount of compression, but visual quality will start to degrade noticeably beyond a certain point.

Video Resolution

The Video Resolution setting specifies the source and target video resolutions. The default setting is **Full HD (1080p) to Full HD (1080p)**.

- Ultra HD (2160p) to Ultra HD (2160p)
- Ultra HD (2160p) to Full HD (1080p)
- Full HD (1080p) to Full HD (1080p)

Test Objective

The Test Objective setting lets you specify the objective of the benchmark. The default setting is **Performance and quality.**

When you choose **Performance and quality** the test measures the number of parallel streams, minimum frame rate, and throughput frame rate to assess performance, and PSNR and bitrate to assess quality.

If you select **Quality**, the test measures only PSNR and bitrate.

GOOD TESTING GUIDE

To ensure accurate and consistent benchmark results you should test clean server systems without third party software installed. If this is not possible, try to close as many background tasks as possible, especially those that might interrupt the benchmark run with updates or notifications.

> ⚠️ Running other programs on the server while the benchmark is running will affect the results.

You can cancel a benchmark run by pressing **CTRL+C** on the Servermark Client system or with the **Cancel** button in the Servermark Controller.

Recommended process

1. Install critical updates and drivers to ensure your server is up to date.
2. Reboot both the Servermark Controller and Client systems.
3. Ensure that background applications are closed or disabled.
4. Verify that the Servermark Client is running.
5. Start the Servermark Controller and add the Client(s).
6. In the Controller, configure the options and run the benchmark.
7. Run the benchmark at least three times to verify your results.

Expert users should read the Video Quality section for a guide to more advanced ways of using the Servermark Media Transcode benchmark.

BENCHMARK WORKLOAD

The Media Transcode benchmark uses royalty-free video clips that are publicly available. The video clips are 60 seconds long and encoded in high-quality H.264 at 30 frames per second. The clips do not have an audio track. To stress the encoder in varied ways, three distinct types of clip are used:

- High motion
- Low motion
- Low light

High- and low-motion video clips differ in the amount of motion vectors a motion-estimation algorithm will find in the video. The amount of motion vectors depends on the encoder and encoder parameters, which limit the compute resources for finding optimal motion vectors.

Motion estimation is a computationally heavy process. Encoding high-motion video tends to be considerably slower than encoding low-motion video. As a result, the low-motion video typically represents the best case while results from the high-motion video show the worst case.

Low-light video is statistically between low- and high-motion video, but it also has the property of featuring dark scenes with details visible in the shadows. The objective of low-light video testing is to make sure that the quality of the encoder output is acceptable in this demanding scenario.

High motion

For the high-motion scene, we use parts of following videos:

- The **ducks_take_off** clip from The SVT High Definition Multi Format test set used under license.
- A five second **downhill biking** clip was extracted from a longer video available from Shutterstock.
- Two five second clips were extracted from a **rollercoaster** video available from Shutterstock.
- The **horses running across a river clip** is available from Shutterstock.

We combined the clips to create a 60-second test video:

- Ducks take off (10 seconds)
- Downhill biking (5 seconds)
- Rollercoaster near ground (5 seconds)
- Horses running across a river (5 seconds)
- Rollercoaster upside down (5 seconds)
- The videos then repeat once more in the same order

Low motion

The low-light scene uses the **old_town_cross** clip from The SVT High Definition Multi Format test set available under license.

The clip is looped six times to make a 60 second test video. Additional noise reduction filtering was applied to make sure that motion estimators are not distracted by camera noise that can be especially present in 4K video.

Low light

For the low-light scene we selected a clip called **Autumn Road 02, National Park of Geres, Portugal** from Shutterstock. The clip was looped to make it 60 seconds long.

SCORING
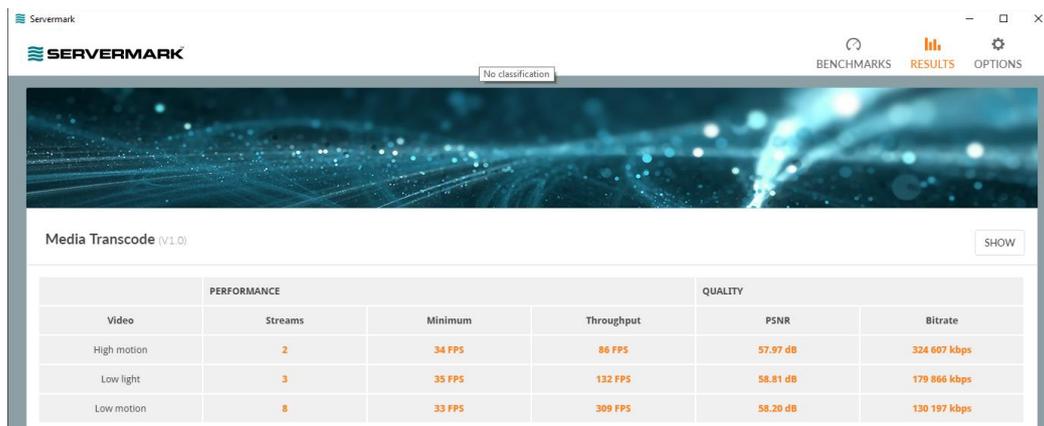
The Media Transcode benchmark produces five different scores for all scenes. These scores are calculated from the final round of search methodology that was used.

- Streams
- Minimum frame rate
- Throughput frame rate
- PSNR
- Bitrate.

Performance scores

If the aim is to measure performance, **Streams** is the most important score. It tells you how many real-time streams the servers are able to stream before the stream is no longer real time. Please read the Transcoding engines section for a more detailed explanation of performance.

When the number of streams are identical, you should then consider **Throughput**, which represents how many frames per second the server was able to stream. The **Minimum** frame rate tells you how close the stream(s) came to falling below real time.

Quality scores

If video quality is the objective of the test, the important scores are **PSNR** and **bitrate**. Please read the Video Quality section for a more detailed explanation of video quality.

The peak-signal-to-noise ratio (**PSNR**) is an image quality metric that is measured in decibels (dB). It is an expression of the ratio of the maximum possible value (power) of the signal and the power of the distorting noise that affects the quality of its representation. High PSNR values indicate good quality, whereas low values indicate poor quality.

Bitrate is the amount of data dedicated to each second of video, measured in kbps. The higher the bitrate, the more bandwidth is needed to play back the video clip.

Streams

Each Servermark Client determines how many streams it is able to stream before the frame rate drops below real time. Servermark calculates all streams together as group streams (S_{group}) as follows:

$$S_{group} = \sum_{i=1}^{N} x_i$$

Where:

x	=	Node streams
N	=	The number of Servermark Clients

Minimum frame rate

The first five seconds are not included in the measurement. Each Client stream first splits the scene into two-second time windows. Next, the Client checks the minimum time window from all streams. Finally, Servermark checks the minimum time window from all Clients to get the group minimum frame rate (S_{min}):

$$S_{min} = min_{1...N}\left(min_{1...M}\left(min_{1...L}(x_i)\right)\right)$$

Where:

x	=	Stream time window
L	=	The number of stream time windows
M	=	The number of streams
N	=	The number of Servermark Clients

Throughput

The first five seconds are not included in the measurement. Each Client stream first calculates the overall frame rate. Then Servermark takes a sum of those frame rates to get the node throughput that the Client was able to produce. Finally, Servermark calculates these node throughputs together to get group throughput ($S_{throughput}$):

$$S_{throughput} = \sum_{i=1}^{N}\left(\sum_{j=1}^{M} x_i\right)$$

Where:

x	=	Stream overall frame rate in frames per second
M	=	The number of streams
N	=	The number of Servermark Clients

PSNR

Each Servermark Client first calculates the PSNR per scene—x_i—by comparing the reference scene with the first scene transcoded. Finally, Servermark checks the minimum PSNR from all PSNR values per scene (S_{psnr}):

$$S_{psnr} = min_{1...N}(x_i)$$

Where:

x	=	Node scene PSNR
N	=	The number of Servermark Clients

Bitrate

Each Servermark Client first calculates bitrate per scene—x_i—then Servermark checks the maximum bitrate value from all bitrates per scene ($S_{bitrate}$):

$$S_{bitrate} = max_{1...N}(x_i)$$

$$x_i = \frac{8 \times filesize}{duration \times 1024}$$

Where:

x	=	Node clip bitrate
N	=	Number of Servermark Clients
$filesize$	=	Size (bytes) of the clip video
$duration$	=	Length (seconds) of the clip = 60

TRANSCODING ENGINE

There are two different implementation using different runtimes. The Intel implementation uses Intel Media Server Studio 2017 runtimes. The NVIDIA implementation uses NVIDIA graphics drivers.

TRANSCODING ENGINE

Intel	Intel Media Server Studio 2017 R2 SDK
NVIDIA	NVIDIA Video Codec SDK 7.1

Each Client performs a binary search to find the maximum number of streams it is able to support before the frame rate drops below real time.

The search runs with one stream, then two, then four, and so on, until there is a stream that is not real time. The search then goes back and forth until the exact stream number is found.

The benchmark runs an extra three verification rounds to verify that it is not an exceptional event.

The Servermark Client uses one decoder to read the scene source clip and then uses multiple encoders to create multiple streams of the scene.

Complete video output files are written to disk only on the first pass of the test when the performance with one video output stream is tested. On later passes, only the SHA-512 checksum of each output file is written to avoid an I/O bottleneck and excessive use of storage space. You can use the checksum files to verify that all test passes produce bitwise-similar output.

VIDEO QUALITY

Along with performance metrics, the benchmark calculates a video quality metric, PSNR, and output video bitrate for comparing video quality. These two numbers can be used as additional information when:

1. benchmarking different systems, possibly from different vendors,
2. changing the Quality Level setting,
3. comparing encoder performance with high- and low-motion video.

For a more in-depth comparison of video encoders, we suggest using BD-RATE or BD-PSNR metrics.

The basic idea is to choose a range of QP values (typically four values like 22, 27, 32, 37 or 27, 32, 37, 42) and then run the benchmark with each of these QP values without changing any other settings.

For each test round, the output bitrate and PSNR are recorded. By plotting the bitrate against PSNR (and fitting a curve) one gets a rate-distortion (RD) curve. By repeating the process for the other encoder (using the same encoding settings), one gets another rate-distortion curve.

These two RD curves can be compared with either a BD-RATE or BD-PSNR metric that measures how far away the curves are from each other.

For more information on BD-RATE and BD-PSNR, please see:

- An implementation of metrics that can be run with Matlab or Octave.
- A more thorough explanation of how the metrics are calculated.


Examples

The chart on the next page shows an example of two RD curves where PSNR is plotted as a function of bitrate. The encoder represented by the orange line is more efficient as it produces the same PSNR with less bitrate.

Low motion, AVC, FHD, Best speed
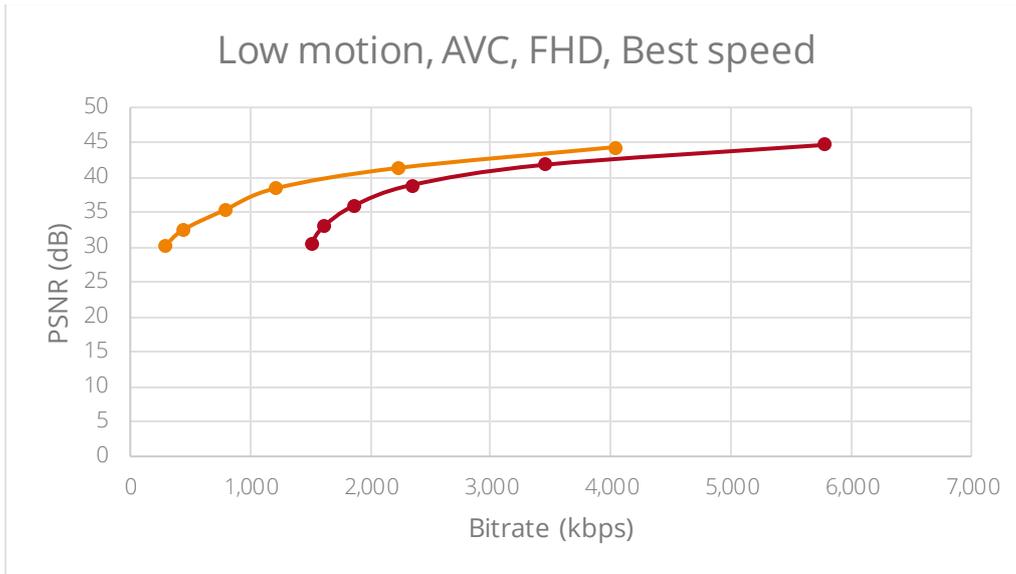
| BD-PSNR | -53.10 |
| BD-RATE | 3.35 |

The second example of RD below is a case where high-motion video was tested. Now the difference of output quality appears to be smaller, which is confirmed by the metrics.

High motion, AVC, FHD, Best speed
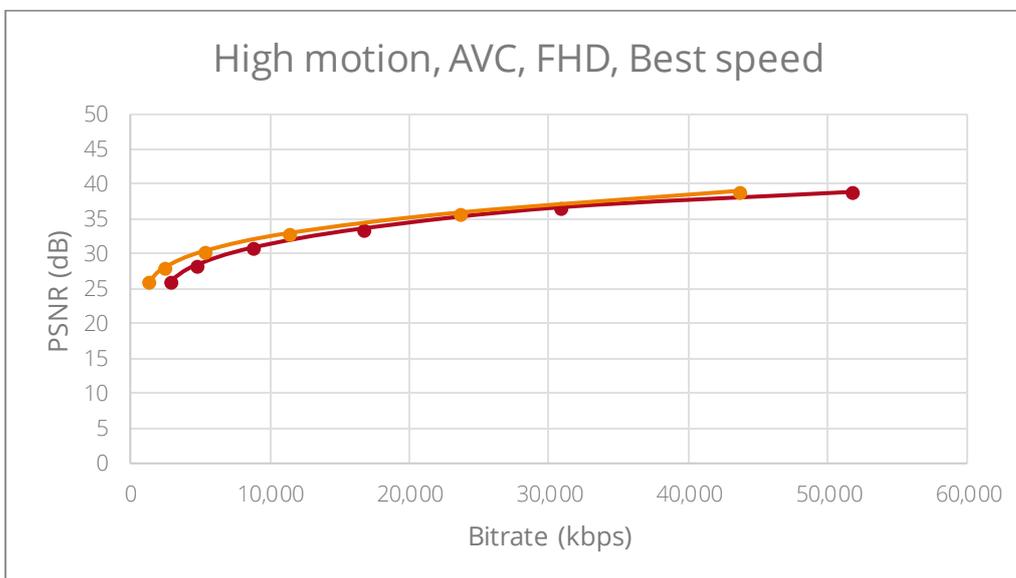
| BD-PSNR | -27.00 |
| BD-RATE | 1.15 |

Video quality is a controversial subject since it is also subjective. The quality metrics discussed above are based on PSNR, which is well accepted as a generic signal-quality metric. But it is far from optimal as an estimator of subjective video quality.

There are other metrics like PSNR-HVS, PSNR-HVS-M, SSIM, MS-SSIM and many more that try account for how the human visual system works.

Still, some experts may rely only on their own visual inspection instead of any computational metric. To enable you to use whatever quality tool or methods you choose after a benchmark run, the video files are left in the Client directory. The input files are located the `test_input` directory and the output files are in the `test_output` directory with informative names.

MEDIA TRANSCODE VERSION HISTORY

VERSION	NOTES
1.0	Launch version

RELEASE NOTES

Servermark v1.0.433 - November 20, 2017

First public release.

New

- Servermark Media Transcode benchmark.
- Added German, Russian, and Simplified Chinese language options.

Servermark v1.0.408 - October 27, 2017

Limited preview release to press.

ABOUT UL

UL is an independent, global company that offers a wide range of testing, inspection, auditing, and certification services. With 10,000 people in 40 countries, UL helps customers, purchasers, and policymakers navigate market risk and complexity. UL builds trust in the safety, security, and sustainability of products, organizations and supply chains – enabling smarter choices and better lives. Visit https://www.ul.com/ to find out more.

UL benchmarking software is developed by the Product Supply Chain Intelligence division. We enable global product compliance, innovation and promotion throughout the supply chain with our intelligent software and services backed by world-class scientific and technical expertise. Please visit https://psi.ul.com/ to find out more.

UL benchmarks help people measure, understand and manage computer hardware performance. Our talented team creates the industry's most trusted and widely used performance tests for desktop computers, notebooks, tablets, smartphones, and VR systems.

We work in cooperation with leading technology companies to develop industry-standard benchmarks that are relevant, accurate, and impartial. As a result, our benchmarks are widely used by the press. UL maintains the world's largest and most comprehensive hardware performance database, using the results submitted by millions of users to drive innovative online solutions designed to help people make informed purchasing decisions.

Our benchmarks are developed in Finland just outside the capital Helsinki. We also have a performance lab and sales office in Silicon Valley and sales representatives in Taiwan.

Press   UL.BenchmarkPress@ul.com
Sales   UL.BenchmarkSales@ul.com
Support   UL.BenchmarkSupport@ul.com