Server performance testing

Working with video content is a highly loaded task that requires the server on which WCS is installed to meet certain requirements.

The minimum server requirements are specified in the documentation, but to determine whether the hardware performance is sufficient for your project, you need to perform a series of load tests according to the conditions of your typical WCS use.

To run the test, you will need the following:

- Virtual (VPS) or physical server
- Video broadcast source (streaming video from OBS studio or ManyCam, broadcasting from your web or IP camera)
- WCS server to emulate viewers of your content

Let’s check the capabilities of Linux x86_64 server with performance that meets the minimum requirements for WCS:

- 2 GB RAM
- 10 GB of disk space
- 1 CPU core

Server Preparation

1. Be sure to specify the size of Java memory heap 1 GB in the WCS server kernel launch configuration file wcs-core.properties:

```
-Xmx1024M
```

After that, restart WCS.

2. All modern server CPUs are multi-core. We will carry out load tests using only one core CPU, disabling all the others. To do this, let's check the current status of the CPU cores used:

```
[root@demo ~]# lscpu | grep list
On-line CPU(s) list: 0-3
```

It appears that there are 4 cores CPU (0,1,2,3) on this server. Disable all cores except the fourth:

```
echo 0 | sudo tee /sys/devices/system/cpu/cpu0/online
echo 0 | sudo tee /sys/devices/system/cpu/cpu1/online
echo 0 | sudo tee /sys/devices/system/cpu/cpu2/online
```

Current status after command execution:

```
[root@demo ~]# lscpu | grep list
On-line CPU(s) list:   3
Off-line CPU(s) list:  0-2
```

For the test, we use 1 CPU Intel Xeon E3-1240 v5@3.50GHz.

Testing

Stream transcoding test

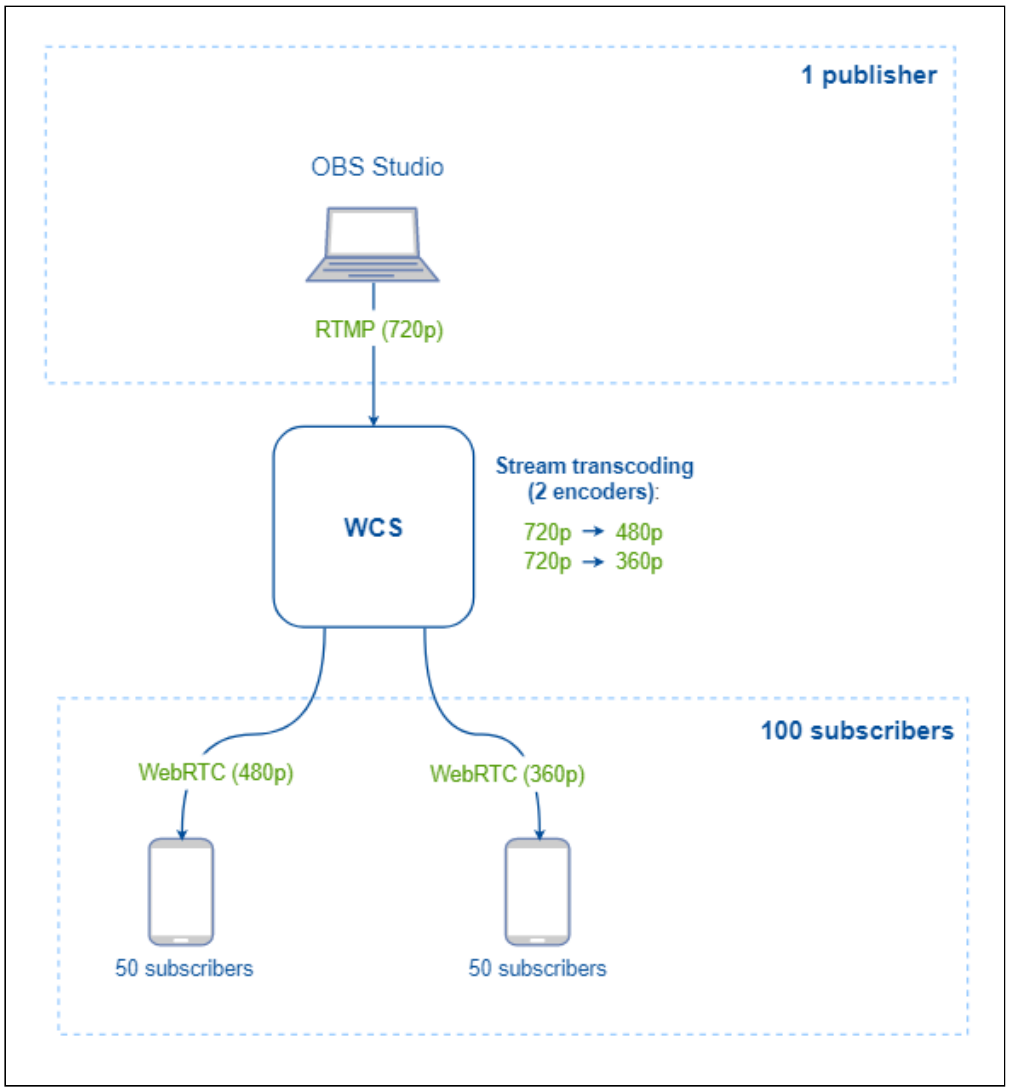
1. We publish an RTMP stream with certain quality parameters from OBS Studio to a WCS server (an example shown in the documentation).

Resolution	Bitrate, kbps
1280x720 (720p)	1500

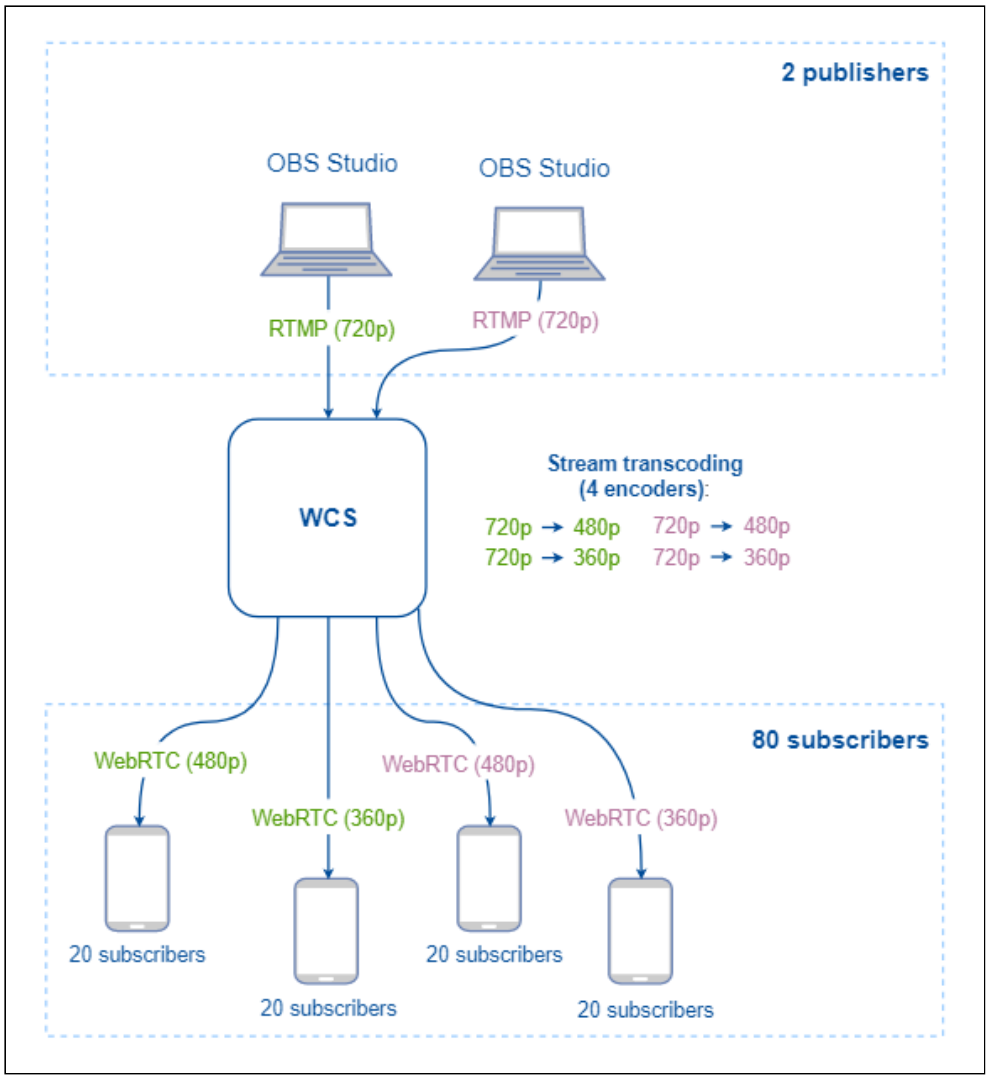
2. We transcode this stream into several popular resolutions (480p and 360p) using REST API.

Resolution	Bitrate, kbps
854x480 (480p)	1000
640x360 (360p)	500

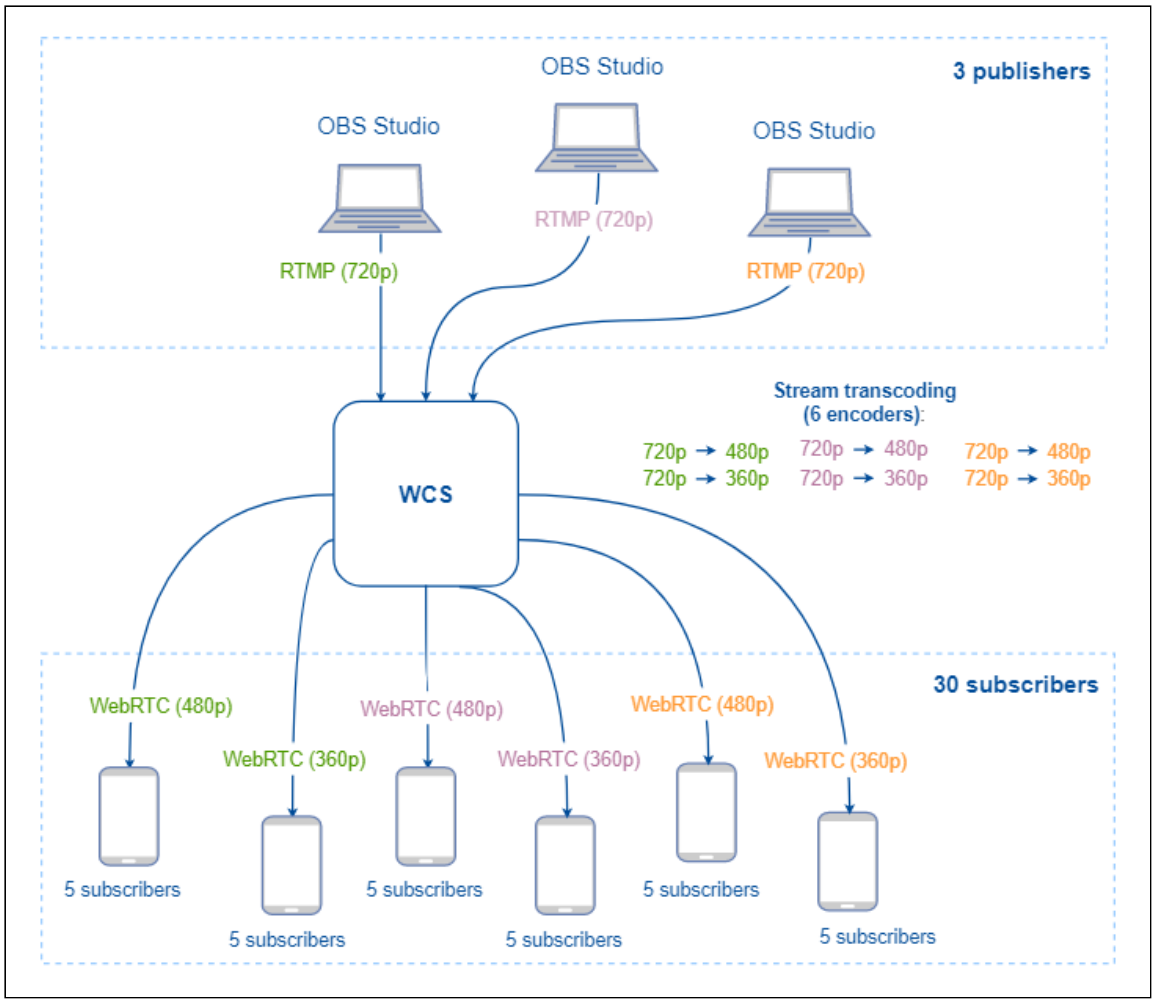
3. We use an example of load testing with capture of streams via WebRTC on another WCS server. With this example, we emulate viewers (subscribers) of a broadcast reproducing the stream in different resolutions (720p, 480p, 360p). At a given number of viewers (about 100), the CPU load on the WCS server is about 80%; this is the recommended maximum CPU load, at which the server functions correctly.



4. We publish two 720p streams, transcode them to 480p and 360p, add them to load testing and determine the maximum number of subscribers (about 80) of these broadcasts according to the processor load on a server with WCS (as described above, its permissible load is up to 80%).



5. Similarly, after doing a test with three 720p streams, we get a possible number of viewers of about 30.

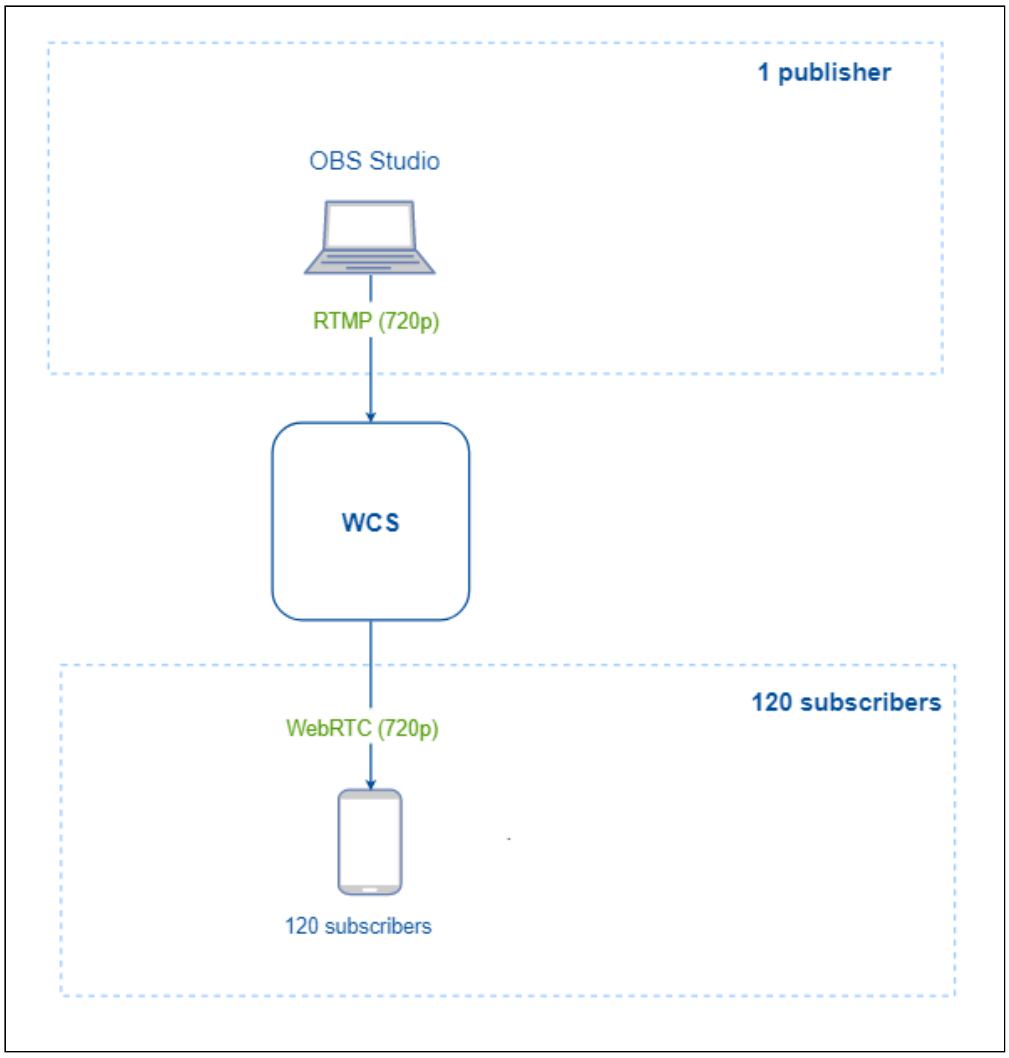


Test without stream transcoding

1. . In this test, we'll check how many viewers will be able to watch the broadcast without transcoding the stream on the server, i.e. we publish an RTMP stream with certain quality parameters from OBS Studio to the WCS server (an example shown in the documentation) and view it as subscribers.

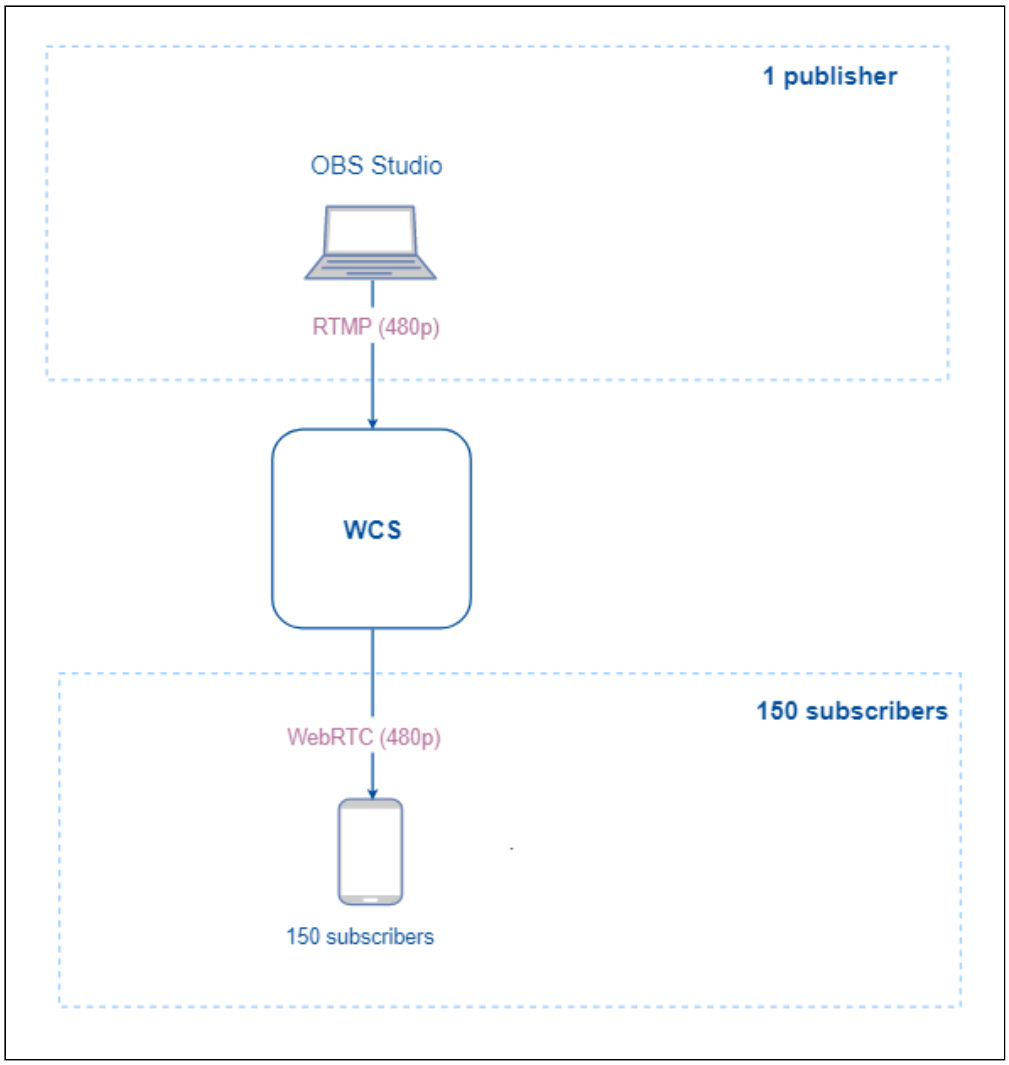
Resolution	Bitrate, kbps
1280x720 (720p)	1500

2. We use an example of load testing with capturing of streams via WebRTC on another WCS server. With this example, we emulate viewers (subscribers) of a broadcast playing a stream from the WCS server. We increase the number of viewers until the processor load parameters on the WCS server reach 80%. When watching a 720p broadcast, we got the possible number of stream viewers – 120 subscribers.



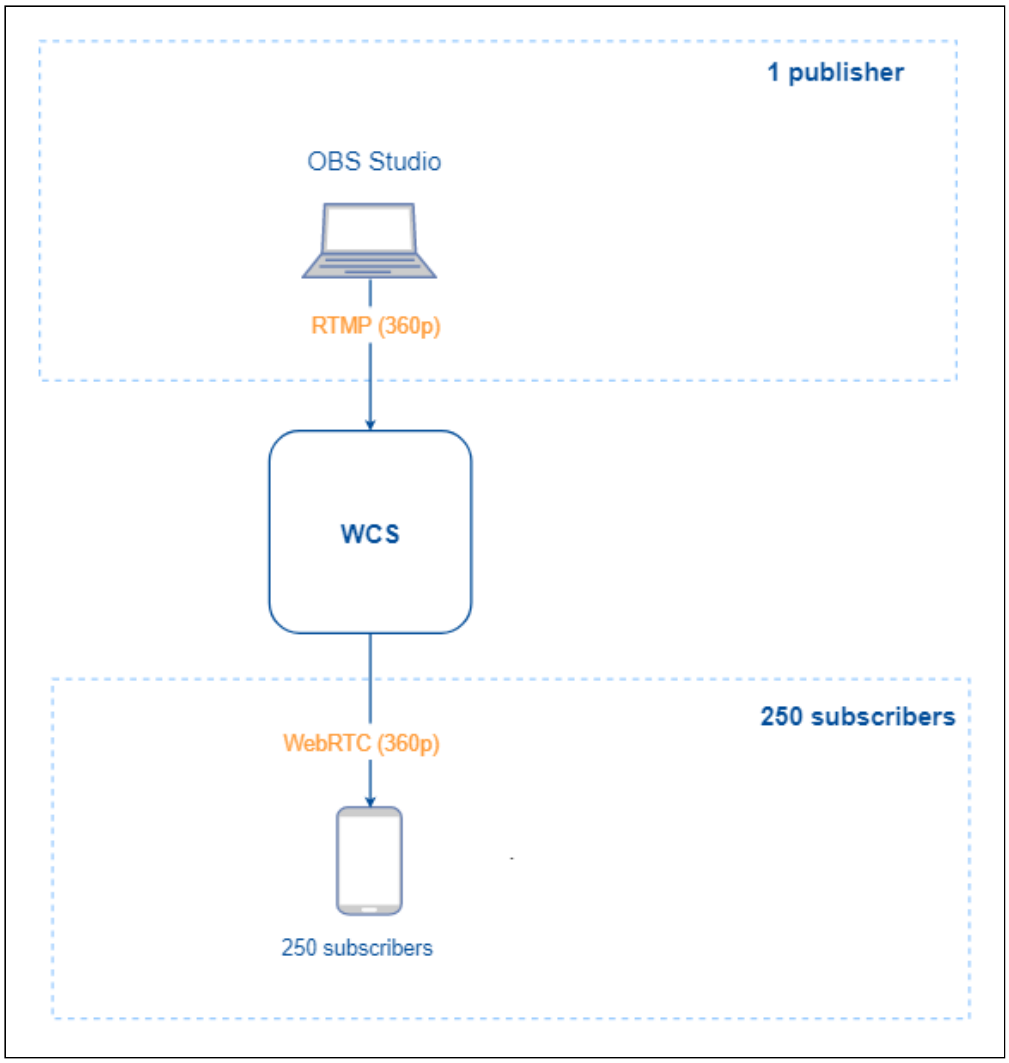
3. We then repeat the test with 480p broadcast and load test with stream capture on another WCS server.

Resolution	Bitrate, kbps
854x480 (480p)	1000



4. Let’s check the capabilities of the server with 360p broadcast and load test.

Resolution	Bitrate, kbps
640x360 (360p)	500



Test Results

Physical server

Based on the test with the minimum recommended configuration (1 CPU, 2 GB RAM, 1 GB RAM for Java heap) on a dedicated (physical) server, we determined the approximate WCS capabilities for working with streaming video on such a server:

Without transcoding

Test

Publications

Viewers

Quantity

Resolution

Bitrate, kbps

Quantity

1

1

1280x720 (720p)

1500

120

2

1

854x480 (480p)

1000

150

3

1

640x360 (360p)

500

250

With transcoding

Test

Publications

Viewers

Quantity

Resolution

Bitrate, kbps

Resolution

Bitrate, kbps

Quantity

1

1

1280x720 (720p)

1500

854x480 (480p)

1000

50

640x360 (360p)

500

50

2

1

1280x720 (720p)

1500

854x480 (480p)

1000

20

640x360 (360p)

500

20

1

1280x720 (720p)

1500

854x480 (480p)

1000

20

640x360 (360p)

500

20

3

1

1280x720 (720p)

1500

854x480 (480p)

1000

5

640x360 (360p)

500

5

1

1280x720 (720p)

1500

854x480 (480p)

1000

5

1

1280x720 (720p)

1500

854x480 (480p)

1000

5

640x360 (360p)

500

5

Virtual Server

A virtual server with similar parameters offers lower performance than a physical server due to a number of reasons (for example, features of a virtualization system), but allows scaling the computing power in the shortest possible time, if necessary.

Minimum recommended configuration test results (1 CPU, 2 Gb RAM, 1 GB RAM for Java heap) on a virtual server by Digital Ocean (digitalocean.com):

Without transcoding

Test

Publications

Viewers

Quantity

Resolution

Bitrate, kbps

Quantity

1

1

1280x720 (720p)

3000

50

2

1

854x480 (480p)

1800

70

3

1

640x360 (360p)

1300

70

With transcoding

Test

Publications

Quantity

Resolution

Bitrate, kbps

1

2

1280x720 (720p)

3000

2

3

854x480 (480p)

1800

3

5

640x360 (360p)

1300

Cloud server (AWS for example)

Test on Amazon cloud servers (virtual and metal) were performed with the following tuning

1. WebRTC traffic encryption hardware acceleration enabled

2. Stream distribution optimization enabled

A stream 1080p with bitrate 2,2 Mbps without bitrate peaks was published to server

The following results are obtained:

Instance type	CPUs	RAM, Gb	Bandwidth, Gbps	Subscribers quantity
c5.4xlarge	16	32	up to 10	1500
c5.9xlarge	36	72	10	2000 (bandwidth limit reached)
c5n.9xlarge	36	96	50	3000

Recommendations

According to the tests, we can conclude that a physical server with similar hardware parameters shows greater performance compared to a virtual server. A variety of devices for viewing and working with streaming video (both mobile platforms and web-integration of content) and limitations on the capacity of network channels available to viewers require significant resources for transcoding streams on the WCS server. Sample server performance requirements for WCS for typical tasks are listed below:

Subscribers quantity	CPUs	RAM, GB	Traffic, TB	Use case
up to 200	4	8	5	CCTV system
up to 500	8	16	6	Webinars
up to 1000	16	64	9	Video conference
up to 2000	20	96	10	HD video streaming

With more streams and viewers and complication of the business model of the project, increasing the productivity of one WCS server is impractical and leads to the emergence of a single point of failure. Scaling, geographical and logical separation (with allocation depending on the performance and role of individual servers in the CDN functions of transcoding and content delivery) allows us to flexibly determine the required level of performance for each WCS server based on the tests we have proposed and performed.

Attachments:

- perf-1.png (image/png)
- perf-2.png (image/png)
- perf-3.png (image/png)
- perf-2-1.png (image/png)
- perf-2-2.png (image/png)
- perf-2-3.png (image/png)