

Настройка балансировки нагрузки с масштабированием при помощи AWS ELB

- [Описание](#)
- [Запуск классического балансировщика нагрузки на базе собственного образа с автоматическим масштабированием](#)
 - 1. Создание образа
 - 2. Создание балансировщика нагрузки
 - 3. Создание новой конфигурации запуска
 - 4. Создание новой группы масштабирования
- [Запуск балансировщика нагрузки для приложений с использованием существующих серверов](#)
 - 1. Запуск серверов
 - 2. Создание балансировщика нагрузки
 - 3. Создание целевой группы для WebSocket порта
 - 4. Редактирование порта балансировщика
- [Запуск балансировщика нагрузки для приложений на базе образа из AWS Marketplace с автоматическим масштабированием](#)
 - 1. Создание шаблона запуска (Launch Template)
 - 2. Создание группы масштабирования
- [Тестирование](#)

Описание

Экземпляры WCS на Amazon поддерживают балансировку нагрузки при помощи AWS elastic load balancer (ELB).

При этом WebSocket-соединения будут автоматически распределены между активными серверами в балансировщике нагрузки. В случае применения заданной политики масштабирования (если целевой показатель, например, загрузка процессора на сервере, достиг заданного значения) будут запущены новые экземпляры сервера и автоматически добавлены в балансировщик.

Для настройки потребуются следующие компоненты

- Образ, на основе которого будут создаваться новые экземпляры сервера при масштабировании
- Балансировщик нагрузки
- Конфигурация для запуска
- Группа масштабирования

Запуск классического балансировщика нагрузки на базе собственного образа с автоматическим масштабированием

Создание балансировщика нагрузки с автомасштабированием на базе собственного образа полезно для длительной эксплуатации (месяцы, годы). В этом случае базовый образ из AWS Marketplace обойдется дороже из-за почасовой оплаты, поэтому рекомендуется приобрести и активировать собственную лицензию WCS.

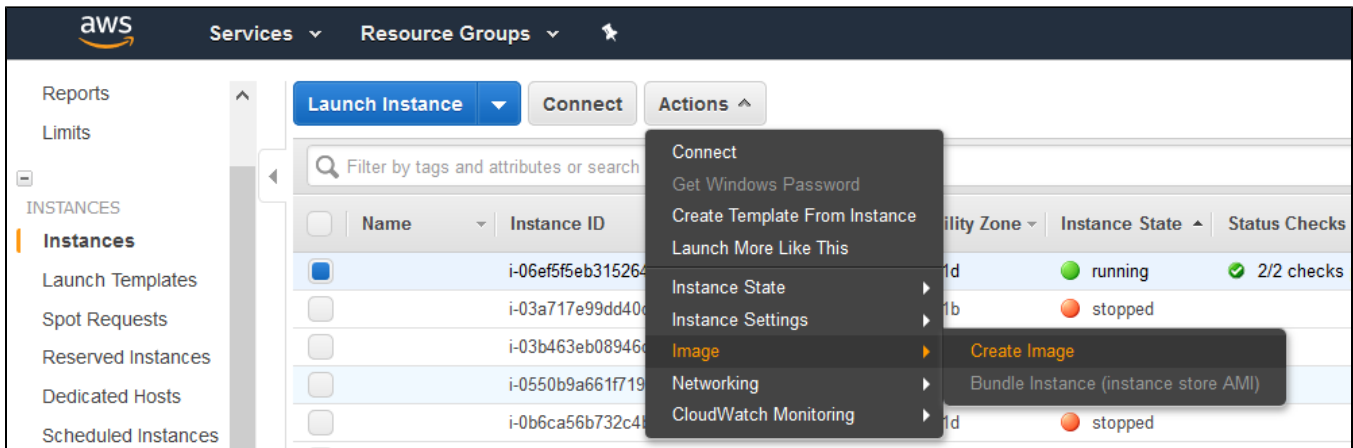
Обратите внимание, что классический балансировщик нагрузки перестанет поддерживаться в августе 2022 года.

1. Создание образа

1.1. [Создайте экземпляр](#) из готового образа FlashphonerWebCallServer AMI и настройте WCS

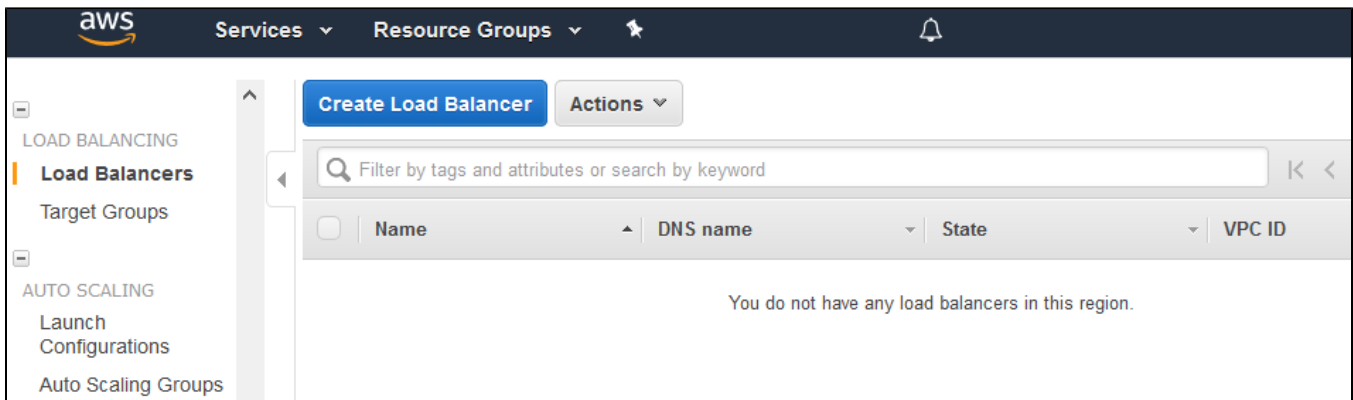
- активируйте лицензию
- импортируйте SSL сертификат
- настройте сервер

1.2. В консоли AWS, выберите созданный экземпляр, затем в меню "Actions" | "Image" | "Create Image" создайте новый образ:

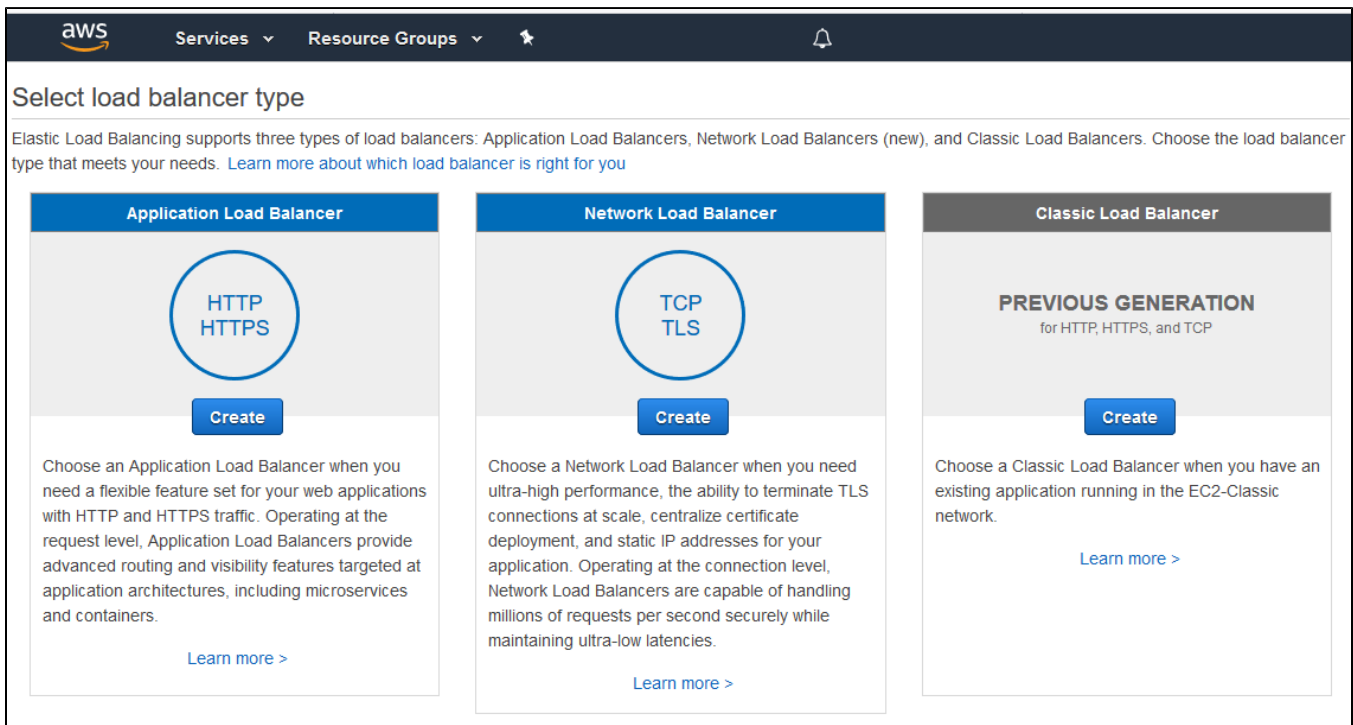


2. Создание балансировщика нагрузки

2.1. В консоли AWS, перейдите в пункт “EC2” | “Load Balancers” и нажмите “Create Load Balancer”



2.2 Выберите тип “Classic Load Balancer” (Данный тип позволяет выбрать порты для контроля состояния сервера)



2.3. На первом шаге настройки, добавьте требуемые протоколы. Например TCP, port 8080 для WebSocket соединений (ws:<host>:8080).

aws

Services

Resource Groups

1. Define Load Balancer

2. Assign Security Groups

3. Configure Security Settings

4. Configure Health Check

5. Add EC2 Instances

6. Add Tags

7. Review

Step 1: Define Load Balancer

Basic Configuration

This wizard will walk you through setting up a new load balancer. Begin by giving your new load balancer a unique name so that you can identify it from other load balancers you might create. You will also need to configure ports and protocols for your load balancer. Traffic from your clients can be routed from any load balancer port to any port on your EC2 instances. By default, we've configured your load balancer with a standard web server on port 80.

Load Balancer name:

TEST-LB

Create LB Inside:

My Default VPC (172.31.0.0/16)

Create an internal load balancer:

☐

(what's this?)

Enable advanced VPC configuration:

☐

Listener Configuration:

Load Balancer Protocol	Load Balancer Port	Instance Protocol	Instance Port	
HTTP	8081	HTTP	8081	✕
TCP	8080	TCP	8080	✕

Add

Cancel

Next: Assign Security Groups

2.4. Определите группу безопасности.

2.5. Настройте проверку состояния сервера

Используйте следующие URL для контроля состояния

- по HTTP: http://WCS_ADDRESS:8081/?action=stat

- по HTTPS:https://WCS_ADDRESS:8444/?action=stat

aws

Services

Resource Groups

1. Define Load Balancer

2. Assign Security Groups

3. Configure Security Settings

4. Configure Health Check

5. Add EC2 Instances

6. Add Tags

7. Review

Step 4: Configure Health Check

Your load balancer will automatically perform health checks on your EC2 instances and only route traffic to instances that pass the health check. If an instance fails the health check, it is automatically removed from the load balancer. Customize the health check to meet your specific needs.

Ping Protocol

HTTP

Ping Port

8081

Ping Path

/?action=stat

Advanced Details

Response Timeout

5

seconds

Interval

30

seconds

Unhealthy threshold

2

Healthy threshold

10

Cancel

Previous

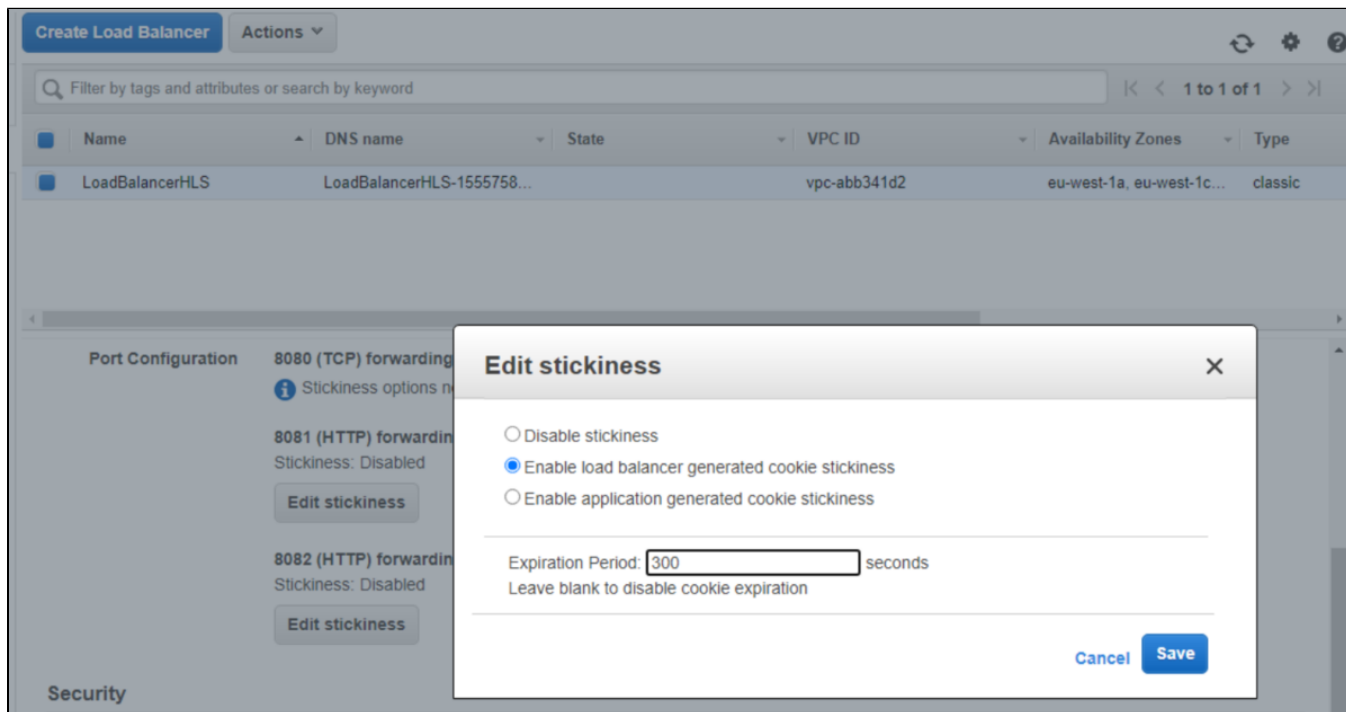
Next: Add EC2 Instances

2.6. Добавьте существующие экземпляры сервера, по необходимости

По умолчанию, балансировщик будет распределять трафик между всеми доступными зонами в Вашем регионе.

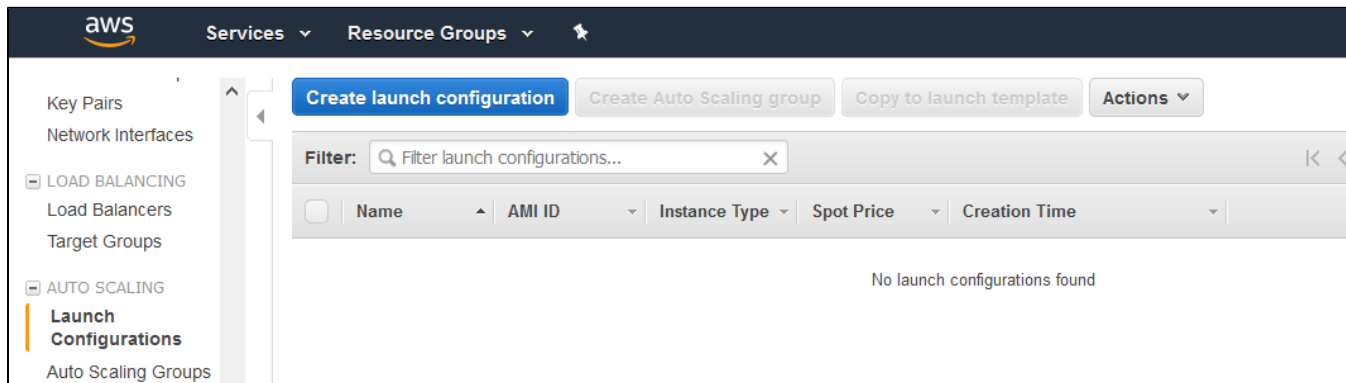
2.7. Завершите создание балансировщика нагрузки

2.8. Включите залипание (stickiness) для HTTP/HTTPS портов балансировщика

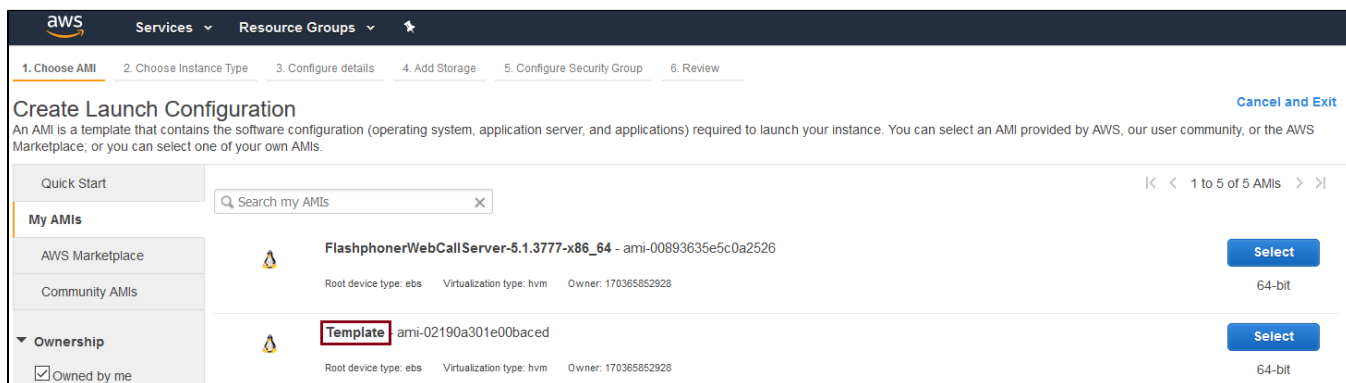


3. Создание новой конфигурации запуска

3.1. В консоли AWS, выберите пункт "EC2" | "Launch Configurations" и нажмите "Create launch configuration"



3.2. Выберите образ, созданный на шаге 1, с требуемой настройкой WCS



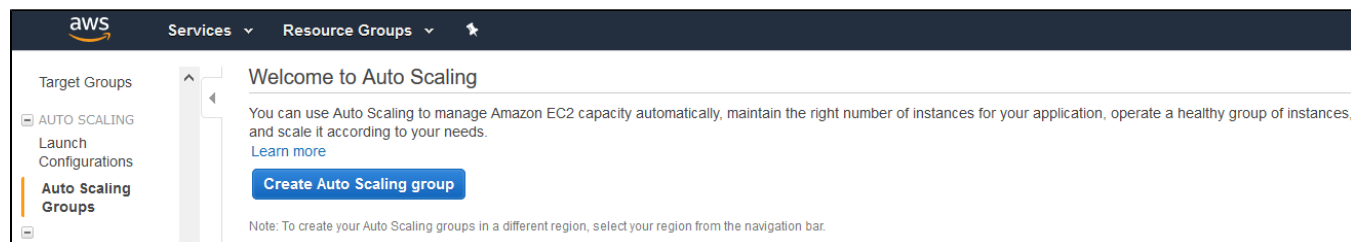
3.3. Завершите создание конфигурации запуска

В разделе Advanced details может быть настроен мониторинг с периодичностью до 1 минуты.

The screenshot shows the AWS Management Console interface for creating a launch configuration. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a star icon. Below the navigation bar is a progress bar with six steps: 1. Choose AMI, 2. Choose Instance Type, 3. Configure details (highlighted), 4. Add Storage, 5. Configure Security Group, and 6. Review. The main heading is 'Create Launch Configuration'. Below it are four configuration sections: 'Name' with a text input field containing 'TEST-LG'; 'Purchasing option' with a checkbox for 'Request Spot Instances' which is unchecked; 'IAM role' with a dropdown menu set to 'None'; and 'Monitoring' with a checked checkbox for 'Enable CloudWatch detailed monitoring' and a 'Learn more' link. Below these sections is an 'Advanced Details' section with a light blue informational box containing a speech bubble icon and the text: 'Later, if you want to use a different launch configuration, you can create a new one and apply it to any Auto Scaling group. Existing launch configurations cannot be edited.' At the bottom right are four buttons: 'Cancel', 'Previous', 'Skip to review' (highlighted in blue), and 'Next: Add Storage'.

4. Создание новой группы масштабирования

4.1. В консоли AWS, выберите пункт "EC2" | "Auto Scaling Groups" и нажмите "Create Auto Scaling group"



4.2. Выберите требуемую конфигурацию запуска, или выберите "Create a new launch configuration" для создания новой

aws

Services ▾

Resource Groups ▾

🔔

Create Auto Scaling Group

Cancel and Exit

Complete this wizard to create your Auto Scaling group. First, choose either a launch configuration or a launch template to specify the parameters that your Auto Scaling group uses to launch instances.

☒ Launch Configuration

You can continue to use your launch configurations if they support the Amazon EC2 features you need. [Learn more](#)

☐ Launch Template New

Launch templates give you the option of launching one type of instance, or a combination of instance types and purchase options. Launch templates include the latest Amazon EC2 features and can be updated and versioned. [Learn more](#)
[Create new launch template](#)

☐ Create a new launch configuration

☒ Use an existing launch configuration

Filter launch configurations... X

< 1 to 1 of 1 Launch Configurations > |

Name	AMI ID	Instance Type	Spot Price	Security Groups
☒ TEST-LG	ami-0cf30f44be371890d	t2.small		sg-03b1bc951522cb94a

Cancel

Next Step

4.3. Настройте детали группы масштабирования

- добавьте требуемые подсети
- добавьте требуемый балансировщик нагрузки

aws

Services ▾

Resource Groups ▾

★

🔔

1. Configure Auto Scaling group details

2. Configure scaling policies

3. Configure Notifications

4. Configure Tags

5. Review

Create Auto Scaling Group

Cancel and Exit

Group name ⓘ

TEST-ASG

Launch Configuration ⓘ

TEST-LG

Group size ⓘ

Start with 1 instances

Network ⓘ

vpc-714e7e15 (172.31.0.0/16) (default)

Create new VPC

Subnet ⓘ

subnet-d3d17ff9(172.31.48.0/20) | Default in us-east-1d

subnet-861169e3(172.31.64.0/20) | Default in us-east-1c

Create new subnet

Each instance in this Auto Scaling group will be assigned a public IP address. ⓘ

▼ Advanced Details

Load Balancing ⓘ

☒ Receive traffic from one or more load balancers

[Learn about Elastic Load Balancing](#)

Classic Load Balancers ⓘ

TEST-LB x

Target Groups ⓘ

Health Check Type ⓘ

☒ ELB ☐ EC2

Health Check Grace Period ⓘ

300 seconds

Monitoring ⓘ

☒ Enable CloudWatch detailed monitoring

[Learn more](#)

Instance Protection ⓘ

Service-Linked Role ⓘ

AWSServiceRoleForAutoScaling

[View Role in IAM](#)

Cancel

Next: Configure scaling policies

4.4. Настройте политики масштабирования

aws

Services

Resource Groups

1. Configure Auto Scaling group details

2. Configure scaling policies

3. Configure Notifications

4. Configure Tags

5. Review

Create Auto Scaling Group

You can optionally add scaling policies if you want to adjust the size (number of instances) of your group automatically. A scaling policy is a set of instructions for making such adjustments in response to an Amazon CloudWatch alarm that you assign to it. In each policy, you can choose to add or remove a specific number of instances or a percentage of the existing group size, or you can set the group to an exact size. When the alarm triggers, it will execute the policy and adjust the size of your group accordingly. [Learn more](#) about scaling policies.

☐ Keep this group at its initial size

☒ Use scaling policies to adjust the capacity of this group

Scale between and instances. These will be the minimum and maximum size of your group.

Scale Group Size

Name:

CPU

Metric type:

Average CPU Utilization

Target value:

70

Instances need:

300

seconds to warm up after scaling

Disable scale-in:

☐

Scale the Auto Scaling group using step or simple scaling policies

Cancel

Previous

Review

Next: Configure Notifications

4.5. Завершите создание группы масштабирования

Запуск балансировщика нагрузки для приложений с использованием существующих серверов

В некоторых случаях набор серверов уже запущен и настроен (например, группа Origin серверов в CDN), и требуется организовать балансировку нагрузки между этими серверами. Для решения этой задачи предназначен Application Load Balancer.

1. Запуск серверов

Запустите и настройте необходимое количество серверов, как описано [здесь](#).

2. Создание балансировщика нагрузки

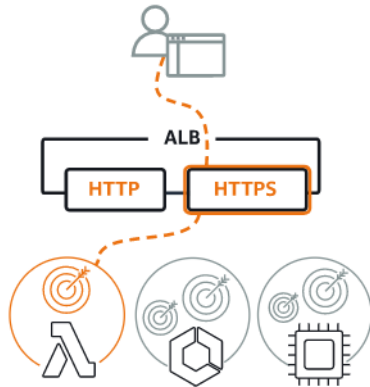
2.1. Перейдите в боковом меню EC2 Console в раздел "Load balancers - Load balancers" и нажмите "Create load balancer". В окне мастера создания балансировщика нажмите Create для Application Load Balancer

Select load balancer type

A complete feature-by-feature comparison along with detailed highlights is also available. [Learn more](#)

Load balancer types

Application Load Balancer [Info](#)

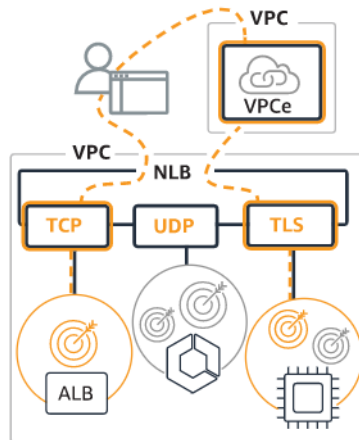


Choose an Application Load Balancer when you need a flexible feature set for your applications with HTTP and HTTPS traffic. Operating at the request level, Application Load Balancers provide advanced routing and visibility features targeted at application architectures, including microservices and containers.

Create



Network Load Balancer [Info](#)



Choose a Network Load Balancer when you need ultra-high performance, TLS offloading at scale, centralized certificate deployment, support for UDP, and static IP addresses for your applications. Operating at the connection level, Network Load Balancers are capable of handling millions of requests per second securely while maintaining ultra-low latencies.

Create

Gateway Load Balancer [Info](#)



Choose a Gateway Load Balancer when you need to deploy and manage a fleet of third-party virtual appliances that support GENEVE. These appliances enable you to improve security, compliance, and policy controls.

Create

2.2. Введите имя балансировщика, выберите тип Internet-facing (предлагается по умолчанию)

Create Application Load Balancer [Info](#)

The Application Load Balancer distributes incoming HTTP and HTTPS traffic across multiple targets such as Amazon EC2 instances, microservices, and containers, based on request attributes. When the load balancer receives a connection request, it evaluates the listener rules in priority order to determine which rule to apply, and if applicable, it selects a target from the target group for the rule action.

► How Application Load Balancers work

Basic configuration

Load balancer name

Name must be unique within your AWS account and cannot be changed after the load balancer is created.

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Scheme [Info](#)

Scheme cannot be changed after the load balancer is created.

☒ Internet-facing

An internet-facing load balancer routes requests from clients over the internet to targets. Requires a public subnet. [Learn more](#)

☐ Internal

An internal load balancer routes requests from clients to targets using private IP addresses.

IP address type [Info](#)

Select the type of IP addresses that your subnets use.

☒ IPv4

Recommended for internal load balancers.

☐ Dualstack

Includes IPv4 and IPv6 addresses.

2.3. В разделе Network mapping выберите необходимые подсети

Network mapping [Info](#)

The load balancer routes traffic to targets in the selected subnets, and in accordance with your IP address settings.

VPC [Info](#)

Select the virtual private cloud (VPC) for your targets. Only VPCs with an internet gateway are enabled for selection. The selected VPC cannot be changed after the load balancer is created. To confirm the VPC for your targets, view your [target groups](#).

-
vpc-e305fc9a
IPv4: 172.31.0.0/16



Mappings [Info](#)

Select at least two Availability Zones and one subnet per zone. The load balancer routes traffic to targets in these Availability Zones only. Availability Zones that are not supported by the load balancer or the VPC are not available for selection. Subnets cannot be removed after the load balancer is created, but additional subnets can be added.

☒ eu-west-1a

Subnet

IPv4 settings

Assigned by AWS

☐ eu-west-1b

☐ eu-west-1c

2.4. Выберите или создайте необходимые security groups

Security groups [Info](#)

A security group is a set of firewall rules that control the traffic to your load balancer.

Security groups

Select security groups

Create new security group [↗](#)

default sg-9c127cdf ✕

VPC: vpc-e305fc9a

Обратите внимание, что в security group для балансировщика должен быть открыт стандартный набор портов WCS (пример)

Create new based on seller settings

A new security group will be generated by AWS Marketplace. It is based on recommended settings for Web Call Server 5 version 5.2.267.

Name your security Group

WCS 5.2

Description

WCS server 5.2

Connection Method	Protocol	Port Range	Source (IP or Group)
SSH	tcp	22	Anywhē 0.0.0.0/0
	tcp	554	Anywhē 0.0.0.0/0
	udp	30000-33000	Anywhē 0.0.0.0/0
	tcp	8080-8084	Anywhē 0.0.0.0/0
	tcp	8443-8445	Anywhē 0.0.0.0/0
	tcp	8888	Anywhē 0.0.0.0/0
	tcp	9091	Anywhē 0.0.0.0/0
	tcp	1935	Anywhē 0.0.0.0/0
	udp	1935	Anywhē 0.0.0.0/0

Rules with source of 0.0.0.0/0 allows all IP addresses to access your instance. We recommend limiting access to only known IP addresses.

Cancel

Save

2.5. В разделе Listeners and routing добавьте Websocket порт (обязательно) и HTTP порт (по необходимости)

Listeners and routing [Info](#)

A listener is a process that checks for connection requests, using the protocol and port you configure. Traffic received by the listener is then routed per your specification. You can specify multiple rules and multiple certificates per listener after the load balancer is created.

▼ Listener HTTP:8080 [Remove](#)

Protocol

HTTP ▼

Port

: 8080

1-65535

Default action [Info](#)

Forward to

test-ws-app-group

Target type: Instance, IPv4

HTTP ▼



[Create target group](#)

▼ Listener HTTP:8081 [Remove](#)

Protocol

HTTP ▼

Port

: 8081

1-65535

Default action [Info](#)

Forward to

test-http-app-group

Target type: Instance, IPv4

HTTP ▼



[Create target group](#)

[Add listener](#)

Для каждого из портов создается целевая группа (target group), см [ниже](#).

2.6. Внизу страницы нажмите Create load balancer

Additional AWS services can be integrated with this load balancer at launch. You can also add these and other services after your load balancer is created by reviewing the "Integrated Services" tab for the selected load balancer.

☐ Create an accelerator to get static IP addresses and improve the performance and availability of your applications. [Additional charges apply](#)

Consider adding tags to your load balancer. Tags enable you to categorize your AWS resources so you can more easily manage them. The 'Key' is required, but 'Value' is optional. For example, you can have Key = production-webserver, or Key = webserver, and Value = production.

Review and confirm your configurations. Estimate cost [↗](#)

Attributes

Cancel Create load balancer

✔ Successfully created load balancer: TEST-APP-LB

Note: It might take a few minutes for your load balancer to be fully set up and ready to route traffic. Targets will also take a few minutes to complete the registration process and pass initial health checks.

EC2 > Load balancers

- Review, customize, or enable attributes for your load balancer and listeners using the **Description** and **Listeners** tabs within **TEST-APP-LB**.
- Discover other services that you can integrate with your load balancer. Visit the **Integrated services** tab within **TEST-APP-LB**.

[View load balancer](#)

3.1. Выберите тип целей Instances (предлагается по умолчанию), задайте имя группы

Step 1

Specify group details

Step 2

Register targets

Specify group details

Your load balancer routes requests to the targets in a target group and performs health checks on the targets.

Basic configuration

Settings in this section cannot be changed after the target group is created.

Choose a target type

☒ **Instances**

- Supports load balancing to instances within a specific VPC.

☐ **IP addresses**

- Supports load balancing to VPC and on-premises resources.
- Facilitates routing to multiple IP addresses and network interfaces on the same instance.
- Offers flexibility with microservice based architectures, simplifying inter-application communication.

☐ **Lambda function**

- Facilitates routing to a single Lambda function.
- Accessible to Application Load Balancers only.

☐ **Application Load Balancer**

- Offers the flexibility for a Network Load Balancer to accept and route TCP requests within a specific VPC.
- Facilitates using static IP addresses and PrivateLink with an Application Load Balancer.

Target group name

test-ws-app-group

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Protocol Port

HTTP ▼ : 8080

3.2. Укажите websocket порт WCS сервера (8080), выберите подсеть и версию протокола (HTTP1)

Protocol Port

HTTP ▼ : 8080

VPC

Select the VPC with the instances that you want to include in the target group.

-
vpc-e305fc9a
IPv4: 172.31.0.0/16 ▼

Protocol version

☒ **HTTP1**
Send requests to targets using HTTP/1.1. Supported when the request protocol is HTTP/1.1 or HTTP/2.

☐ **HTTP2**
Send requests to targets using HTTP/2. Supported when the request protocol is HTTP/2 or gRPC, but gRPC-specific features are not available.

☐ **gRPC**
Send requests to targets using gRPC. Supported when the request protocol is gRPC.

3.3. В разделе Health check настройте проверку работоспособности сервера, используя HTTP порт (8081) и запрос страницы статистики /? action=stat

Health checks

The associated load balancer periodically sends requests, per the settings below, to the registered targets to test their status.

Health check protocol

HTTP

Health check path

Use the default path of "/" to ping the root, or specify a custom path if preferred.

/?action=stat

Up to 1024 characters allowed.

▼ Advanced health check settings

Restore defaults

Port

The port the load balancer uses when performing health checks on targets. The default is the port on which each target receives traffic from the load balancer, but you can specify a different port.

☐ Traffic port

☒ Override

8081

1-65535

Healthy threshold

The number of consecutive health checks successes required before considering an unhealthy target healthy.

5

2-10

Unhealthy threshold

The number of consecutive health check failures required before considering a target unhealthy.

2

2-10

Timeout

The amount of time, in seconds, during which no response means a failed health check.

5

seconds

2-120

и нажмите Next

Interval

The approximate amount of time between health checks of an individual target

30

seconds

5-300

Success codes

The HTTP codes to use when checking for a successful response from a target. You can specify multiple values (for example, "200,202") or a range of values (for example, "200-299").

200

► Tags - optional

Consider adding tags to your target group. Tags enable you to categorize your AWS resources so you can more easily manage them.

Cancel

Next

3.4. На странице Register targets выделите нужные серверы и нажмите Include as pending below

Register targets

Select instances, specify ports, and add the instances to the list of pending targets. Repeat to add additional combinations of instances and ports to the list of pending targets. Once you are satisfied with your selections, click Register pending targets.

Available instances (2/2)

Q Filter resources by property or value

< 1 > ⓘ

☒	Instance ID	Name	State	Security groups	Zone	IPv4 address	Subnet ID
☒	i-Odec078d94e7520ef		running	Web Call Server 5-5-2-944-systemd246-AutogenByAWSMP-	eu-west-1b	3.249.98.141	subnet-41072d27
☒	i-Odbaf422e637b2d9a		running	Web Call Server 5-5-2-944-systemd246-AutogenByAWSMP-	eu-west-1b	34.240.11.186	subnet-41072d27

2 selected

Ports for the selected instances

Ports for routing traffic to the selected instances.

8080

1-65535 (separate multiple ports with commas)

Include as pending below

Затем нажмите Register pending targets

Review targets

Targets (2)

All

Q Filter resources by property or value

< 1 > ⓘ

Remove

Health status

Instance ID

Name

Port

State

Security groups

Zone

IPv4 address

Subnet ID

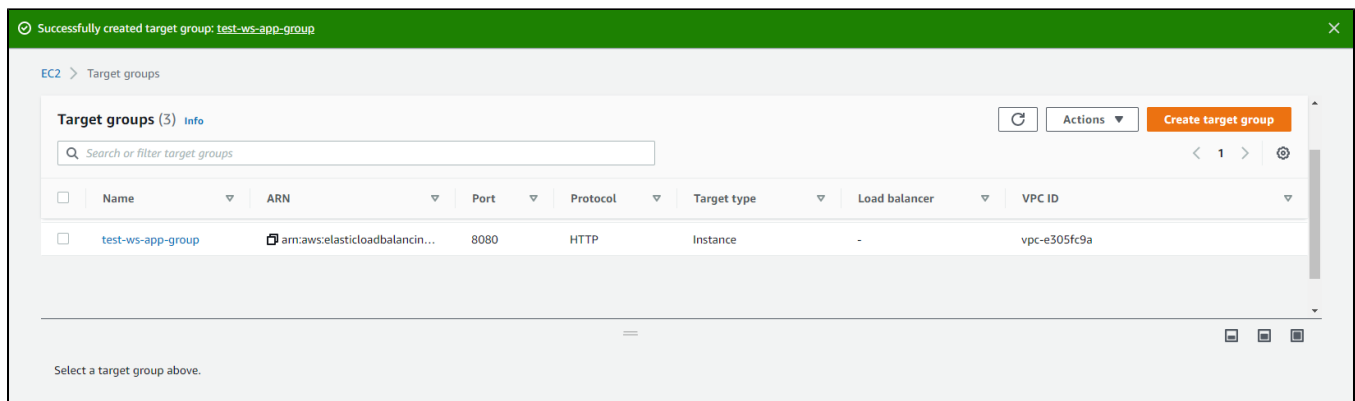
×	Pending	i-Odbaf422e637b2d9a		8080	running	Web Call Server 5-5-2-944-systemd246-AutogenByAWSMP-	eu-west-1b	34.240.11.186	subnet-41072d27
×	Pending	i-Odec078d94e7520ef		8080	running	Web Call Server 5-5-2-944-systemd246-AutogenByAWSMP-	eu-west-1b	3.249.98.141	subnet-41072d27

2 pending

Cancel

Register pending targets

На этом создание целевой группы завершено

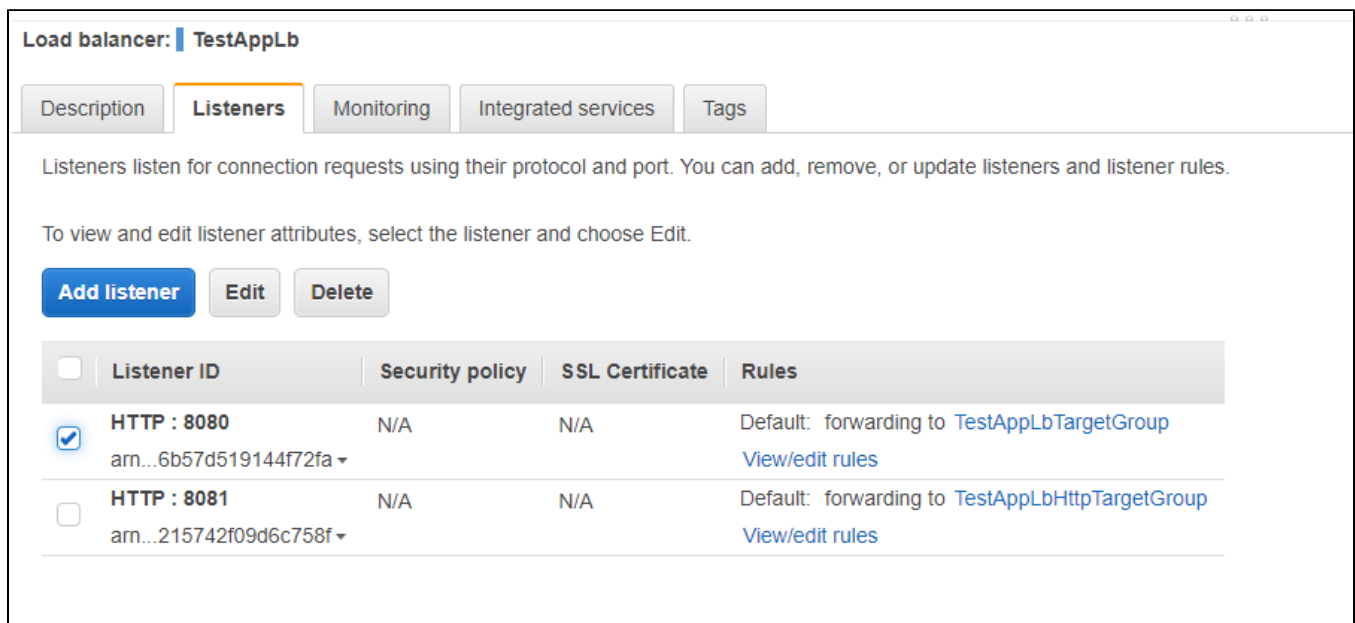


Балансировщик нагрузки, использующий данную группу, начнет работать после того, как хотя бы один сервер в группе успешно пройдет проверку работоспособности.

4. Редактирование порта балансировщика

Если балансировщик создается для группы автоматического масштабирования (см ниже), при его создании нельзя выбрать порт HTTPS, только HTTP. В этом случае настройки порта следует изменить.

4.1. В разделе "Load balancers - Load balancers" выберите вкладку "Listeners" для нужного балансировщика. Выберите Websocket порт и нажмите Edit



4.2. Выберите протокол HTTPS и задайте Secure Websocket порт (например, 8443)

Edit listener

 `arn:aws:elasticloadbalancing:eu-north-1:170365852928:listener/app/TestAppLb/8e650022f70c3d2f/6b57d519144f72fa`

Listener details

A listener is a process that checks for connection requests, using the protocol and port you configure. Traffic received by the listener is then routed per your specification. You can specify multiple rules and multiple certificates per listener after the load balancer is created.

Protocol

HTTPS ▼

Port

8443

1-65535

Default actions [Info](#)

Specify the default actions for traffic on this listener. Default actions apply to traffic that does not meet the conditions of rules on your listener. Rules can be configured after the listener is created.

▼ 1. Forward to [Info](#)

Remove

Target group



Weight (0-999)

TestAppLbTargetGroup

Target type: Instance, IPv4

HTTP ▼

1



Traffic distribution:

100%

Select a target group ▼

0



[Create target group](#) 

☐ [Enable group-level stickiness](#) [Info](#)

If you enable stickiness for your target group, requests routed to it remain in the same group for the duration you specify.

Add action ▼

4.3. В разделе "Secure listener settings" выберите SSL-сертификат из имеющихся для использования для домена, который будет настроен для точки входа балансировщика, или создайте новый. Затем нажмите Save changes

Secure listener settings [Info](#)

Security policy

Your load balancer uses a Secure Socket Layer (SSL) negotiation configuration, known as a security policy, to negotiate SSL connections with clients.

ELBSecurityPolicy-2016-08 ▼

[Compare security policies](#)

Default SSL certificate

The certificate used if a client connects without SNI protocol, or if there are no matching certificates. You can add more certificates after you create the load balancer.

From ACM ▼

*.flashphoner.com
d62e2c7d-23a5-4ef2-8244-6b7dc92d9246 ▼



[Request new ACM certificate](#)

Cancel

Save changes

Настройки порта балансировщика изменены и применяются немедленно

Edit listener

 Suggested next steps

- Review or customize your listener. [Edit listener](#)

View listeners

Запуск балансировщика нагрузки для приложений на базе образа из AWS Marketplace с автоматическим масштабированием

Создание балансировщика нагрузки с автомасштабированием на базе образа из AWS Marketplace полезно для периодического запуска группы серверов, например, при проведении мероприятия (длительностью в часы, дни, недели). В этом случае собственная лицензия с ежемесячной оплатой обойдется дороже, поэтому на короткие промежутки времени рекомендуется использовать образ из Marketplace.

1. Создание шаблона запуска (Launch Template)

1.1. Перейдите в боковом меню EC2 Console в раздел "Instances - Launch Templates" и нажмите "Create launch template". Откроется мастер создания шаблонов запуска. Введите имя и описание шаблона.

Create launch template

Creating a launch template allows you to create a saved instance configuration that can be reused, shared and launched at a later time. Templates can have multiple versions.

Launch template name and description

Launch template name - *required*

Must be unique to this account. Max 128 chars. No spaces or special characters like '&', '*', '@'.

Template version description

Max 255 chars

Auto Scaling guidance [Info](#)

Select this if you intend to use this template with EC2 Auto Scaling

☐ Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

► Template tags

► Source template

1.2. Выберите последний образ FlashphonerWebCallServer

Launch template contents

Specify the details of your launch template below. Leaving a field blank will result in the field not being included in the launch template.

Amazon machine image (AMI) [Info](#)

AMI

FlashphonerWebCallServer-5.2.629-x86_64-hourly-01e37234-6170-4b8d-98b...

ami-035ddfe555ad2e6f8

Catalog: AWS Marketplace architecture: 64-bit (x86) virtualization: hvm

1.3. Выберите тип ВМ, пару ключей для доступа к экземпляру по SSH, группу безопасности

Instance type [Info](#)

Instance type

t2.micro

Free tier eligible

[Instance types](#) 

Family: General purpose 1 vCPU 1 GiB Memory

On-Demand Linux pricing: 0.0126 USD per Hour

On-Demand Windows pricing: 0.0172 USD per Hour

Key pair (login) [Info](#)

Key pair name

test_userdata



[Create new key pair](#)



Network settings

Networking platform [Info](#)



Virtual Private Cloud (VPC)

Launch into a virtual network in your own logically isolated area within the AWS cloud



EC2-Classic

Launch into a single flat network that you share with other customers

Security groups [Info](#)

Select security groups



WCS sg-0ec50e70028ff86d7 ✕

VPC: vpc-e305fc9a

1.4. Задайте размер и параметры диска для экземпляров, создаваемых на базе этого шаблона

Storage (volumes) [Info](#)

▼ Volume 1 (AMI Root)

AMI Volumes are not included in the template unless modified

Volume type [Info](#)

EBS

Device name - required [Info](#)

/dev/sda1

Snapshot [Info](#)

snap-0fee0446fee3252a5

Size (GiB) [Info](#)

10

Volume type [Info](#)

General purpose SSD (gp2)

IOPS [Info](#)

2000

Delete on termination [Info](#)

Yes

Encrypted [Info](#)

No

Key [Info](#)

MyKey

Add new volume

1.5. Разверните раздел формы "Advanced details". В поле "User data" введите скрипт для обновления и настройки WCS

User data [Info](#)

```
#!/bin/bash

# Stop WCS before reconfiguring
PID="$(pgrep -f 'com.flashphoner.server.Server' | grep -v bash)"
if [ -n "$PID" ]; then
    service webcallserver stop
fi

# Update WCS to the latest build (optionally, set to false if you don't)
UPDATE=true
if $UPDATE; then
    cd /tmp
```

☐ User data has already been base64 encoded

Cancel

Create template version

Пример скрипта обновления и настройки Edge сервера для масштабируемой CDN, предназначенной для доставки WebRTC

Edge setup script

```
#!/bin/bash

# Stop WCS before reconfiguring
PID="$(pgrep -f 'com.flashphoner.server.Server' | grep -v bash)"
if [ -n "$PID" ]; then
    service webcallserver stop
fi

# Update WCS to the latest build (optionally, set to false if you don't)
UPDATE=true
if $UPDATE; then
    cd /tmp
    wget --timeout=10 --no-check-certificate https://flashphoner.com/download-wcs5.2-server.tar.gz -O wcs5-server.tar.gz
    if [ $? -eq 0 ]; then
        mkdir -p FlashphonerWebCallServer-5.2-latest && tar xzf wcs5-server.tar.gz -C FlashphonerWebCallServer-5.2-latest --strip-components 1
        cd FlashphonerWebCallServer-5.2-latest
        chmod +x install.sh
        ./install.sh -silent
        cd ..
        rm -rf FlashphonerWebCallServer-5.2-latest wcs5-server.tar.gz
    fi
fi

# Configuration setup
WCS_CONFIG=/usr/local/FlashphonerWebCallServer/conf/flashphoner.properties
JVM_CONFIG=/usr/local/FlashphonerWebCallServer/conf/wcs-core.properties
USERS_CONFIG=/usr/local/FlashphonerWebCallServer/conf/database.yml

#CDN settings
CDN_ROLE=edge
CDN_IP=0.0.0.0
CDN_POINT_OF_ENTRY=172.31.43.82
echo -e "\ncdn_enabled=true" >> $WCS_CONFIG
echo -e "\ncdn_ip=$CDN_IP" >> $WCS_CONFIG
echo -e "\ncdn_role=$CDN_ROLE" >> $WCS_CONFIG
echo -e "\ncdn_point_of_entry=$CDN_POINT_OF_ENTRY" >> $WCS_CONFIG
echo -e "\ncdn_nodes_resolve_ip=false" >> $WCS_CONFIG

# Configure heap settings
HEAP_SIZE=512m
sed -i -e "s/^\(-Xmx\).*$/\1$HEAP_SIZE/" $JVM_CONFIG

# Disable demo user (optionally, set to true if you want to disable)
DISABLE_DEMO=false
if $DISABLE_DEMO; then
    if grep "demo:" $USERS_CONFIG > /dev/null 2>&1; then
        sed -i -e "/demo:/s/active:\ true/active:\ false/" $USERS_CONFIG
    fi
fi

# Start WCS after reconfiguring
PID="$(pgrep -f 'com.flashphoner.server.Server' | grep -v bash)"
if [ -n "$PID" ]; then
    service webcallserver restart
else
    service webcallserver start
fi

# Disable internal firewall, ports are allowed/blocked on security group level
iptables -F
```

Resource tags [Info](#)

No resource tags are currently included in this template. Add a resource tag to include it in the launch template.

Add tag

50 remaining (Up to 50 tags maximum)

Network interfaces [Info](#)

No network interfaces are currently included in this template. Add a network interface to include it in the launch template.

Add network interface

► Advanced details [Info](#)

Cancel

Create launch template

Шаблон запуска будет создан

2. Создание группы масштабирования

EC2 > Auto Scaling groups > Create Auto Scaling group

Step 1
Choose launch template or configuration

Step 2
Configure settings

Step 3 (optional)
Configure advanced options

Step 4 (optional)
Configure group size and scaling policies

Step 5 (optional)
Add notifications

Step 6 (optional)
Add tags

Step 7
Review

Choose launch template or configuration [Info](#)

Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group. If you currently use launch configurations, you might consider migrating to launch templates.

Name

Auto Scaling group name

Enter a name to identify the group.

TestAutoscalingGroup

Must be unique to this account in the current Region and no more than 255 characters.

Launch template [Info](#)

[Switch to launch configuration](#)

Launch template

Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.

TestTemplate

[Create a launch template](#)

Version

Default (1)

[Create a launch template version](#)

Description

Test autoscaling launch template

Launch template

[TestTemplate](#)
lt-04d44d426947bae18

Instance type

-

AMI ID

ami-035ddfe555ad2e6f8

Security groups

-

Security group IDs

-

Key pair name

2.2. Выберите шаблон запуска, укажите версию "Latest"

Launch template [Info](#)

[Switch to launch configuration](#)

Launch template

Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.

TestTemplate ▼



[Create a launch template](#)

Version

Latest (2) ▼



[Create a launch template version](#)

Description

-

Launch template

TestTemplate [↗](#)

lt-04d44d426947bae18

Instance type

t2.micro

AMI ID

ami-035ddfe555ad2e6f8

Security groups

-

Security group IDs

sg-0ec50e70028ff86d7 [↗](#)

Key pair name

test_userdata

Additional details

Storage (volumes)

/dev/sda1

Date created

Tue Jul 14 2020 14:02:03 GMT+0700
(Novosibirsk Standard Time)

2.3. Укажите соотношение между используемыми ВМ (по требованию/с аукциона). По умолчанию предлагается 70 % по требованию, рекомендуется указать 100 %

EC2 > Auto Scaling groups > Create Auto Scaling group

Step 1
Choose launch template or configuration

Step 2
Configure settings

Step 3 (optional)
Configure advanced options

Step 4 (optional)
Configure group size and scaling policies

Step 5 (optional)
Add notifications

Step 6 (optional)
Add tags

Step 7
Review

Configure settings [Info](#)

Configure the settings below. Depending on whether you chose a launch template, these settings may include options to help you make optimal use of EC2 resources.

Purchase options and instance types [Info](#)

☐ Adhere to launch template
The launch template determines the purchase option (On-Demand or Spot) and instance type.

☒ Combine purchase options and instance types
Specify how much On-Demand and Spot capacity to launch and multiple instance types (optional). This choice is most helpful for optimizing the scale and cost for a fleet of instances.

Instances distribution

Optional On-Demand base
Specify how much On-Demand capacity the Auto Scaling group should have for its base portion. The maximum group size will be increased (but not decreased) to this value.

On-Demand Instances

On-Demand percentage above base
Define the percentage split of On-Demand Instances and Spot Instances for your additional capacity beyond the base portion.

% On-Demand

% Spot

Spot allocation strategy per Availability Zone

☒ Capacity optimized - Launch Spot Instances optimally based on the available Spot capacity (recommended)

☐ Lowest price - Launch Spot Instances from the lowest priced instance pools

Number of lowest priced Spot Instance pools to diversify across

Value must be between 1 and 20

2.4 Выберите типы создаваемых ВМ

Instance types [Info](#)

Choose the instance types that best suit the needs of your application.

Primary instance type

Weight [Info](#)

1.

t2.micro
1vCPU 1 Gib Memory

Your launch template does not specify an instance type. As a result, Adhere to launch template cannot be chosen. You can continue by adding an instance type above.

Additional instance types

[Redo recommendations](#)

2.

t2.micro
1vCPU 1 Gib Memory

Add instance type

2.5. Выберите сеть и подсети для создаваемых ВМ

Network Info

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

VPC

vpc-e305fc9a
172.31.0.0/16 Default

↻

[Create a VPC](#)

Subnets

Select subnets

↻

[Create a subnet](#)

Cancel

Previous

Skip to review

Next

2.6. Выберите "Attach to a new load balancer"

Step 1
Choose launch template or configuration

Step 2
Choose instance launch options

Step 3 (optional)
Configure advanced options

Step 4 (optional)
Configure group size and scaling policies

Step 5 (optional)
Add notifications

Step 6 (optional)
Add tags

Step 7
Review

Configure advanced options Info

Choose a load balancer to distribute incoming traffic for your application across instances to make it more reliable and easily scalable. You can also set options that give you more control over health check replacements and monitoring.

Load balancing - optional Info

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

☐ No load balancer
Traffic to your Auto Scaling group will not be fronted by a load balancer.

☐ Attach to an existing load balancer
Choose from your existing load balancers.

☒ Attach to a new load balancer
Quickly create a basic load balancer to attach to your Auto Scaling group.

Attach to a new load balancer

Define a new load balancer to create for attachment to this Auto Scaling group.

Load balancer type
Choose from the load balancer types offered below. Type selection cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, [visit the Load Balancing console](#).

☒ Application Load Balancer
HTTP, HTTPS

☐ Network Load Balancer
TCP, UDP, TLS

2.7. Выберите тип балансировщика "Application Load balancer", задайте имя, выберите Internet-facing, укажите зоны доступности и соответствующие подсети

Attach to a new load balancer

Define a new load balancer to create for attachment to this Auto Scaling group.

Load balancer type

Choose from the load balancer types offered below. Type selection cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, [visit the Load Balancing console](#). 

☒ **Application Load Balancer**
HTTP, HTTPS

☐ **Network Load Balancer**
TCP, UDP, TLS

Load balancer name

Name cannot be changed after the load balancer is created.

TestAppLb

Load balancer scheme

Scheme cannot be changed after the load balancer is created.

☐ Internal

☒ **Internet-facing**

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group. You can select different subnets and add subnets from additional Availability Zones.

VPC

vpc-5e65c237 

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS resolution.

☒ eu-north-1a

subnet-d2cb6fbb ▼

☐ eu-north-1c

Select a subnet ▼

☐ eu-north-1b

Select a subnet ▼

2.8. В разделе "Listeners and routing" укажите WebSocket порт (8080), выберите "Create a target group" и укажите имя целевой группы, которая будет создана

Listeners and routing

If you require secure listeners, or multiple listeners, you can configure them from the [Load Balancing console](#) after your load balancer is created.

Protocol

HTTP

Port

8080

Default routing (forward to)

Create a target group

New target group name

An instance target group with default settings will be created.

TestAppLbTargetGroup

Tags - optional

Consider adding tags to your load balancer. Tags enable you to categorize your AWS resources so you can more easily manage them.

Add tag

50 remaining

Затем нажмите Next

Additional settings - optional

Monitoring [Info](#)

☐ Enable group metrics collection within CloudWatch

Cancel

Previous

Skip to review

Next

2.9. Задайте максимальный размер группы

EC2 > Auto Scaling groups > Create Auto Scaling group

Step 1
Choose launch template or configuration

Step 2
Configure settings

Step 3 (optional)
Configure advanced options

Step 4 (optional)
Configure group size and scaling policies

Step 5 (optional)
Add notifications

Step 6 (optional)
Add tags

Step 7
Review

Configure group size and scaling policies [Info](#)

Set the desired, minimum, and maximum capacity of your Auto Scaling group. You can optionally add a scaling policy to dynamically scale the number of instances in the group.

Group size - *optional* [Info](#)

Specify the size of the Auto Scaling group by changing the desired capacity. You can also specify minimum and maximum capacity limits. Your desired capacity must be within the limit range.

Desired capacity

1

Minimum capacity

1

Maximum capacity

3

Scaling policies - *optional*

Choose whether to use a scaling policy to dynamically resize your Auto Scaling group to meet changes in demand. [Info](#)

☒ **Target tracking scaling policy**
Choose a desired outcome and leave it to the scaling policy to add and remove capacity as needed to achieve that outcome.

☐ None

Scaling policy name

Target Tracking Policy

2.10. Укажите политику масштабирования по загрузке процессора, задайте время запуска экземпляра

Scaling policies - *optional*

Choose whether to use a scaling policy to dynamically resize your Auto Scaling group to meet changes in demand. [Info](#)

- ☒ **Target tracking scaling policy**
Choose a desired outcome and leave it to the scaling policy to add and remove capacity as needed to achieve that outcome.

☐ None

Scaling policy name

Target Tracking Policy

Metric type

Average CPU utilization ▼

Target value

80

Instances need

60

seconds warm up before including in metric

☐ Disable scale in to create only a scale-out policy

Instance scale-in protection - *optional*

Instance scale-in protection

If protect from scale in is enabled, newly launched instances will be protected from scale in by default.

☐ Enable instance scale-in protection

Cancel

Previous

Skip to review

Next

EC2 > Auto Scaling groups > Create Auto Scaling group

Step 1
Choose launch template or configuration

Step 2
Configure settings

Step 3 (optional)
Configure advanced options

Step 4 (optional)
Configure group size and scaling policies

Step 5 (optional)
Add notifications

Step 6 (optional)
Add tags

Step 7
Review

Review Info

Step 1: Choose launch template or configuration Edit

Group details

Auto Scaling group name

TestAutoscalingGroup

Launch template

Launch template	Version	Description
TestTemplate	Latest	
lt-04d44d426947bae18		

Step 2: Configure settings Edit

Purchase options and instance types

Instances distribution

On-Demand base	On-Demand and Spot percentages	Spot allocation strategy
Designate the first 0 instances as On-Demand	100 % On-Demand 0 % Spot	Capacity optimized

Instance types

Instance type	vCPUs	Memory	Network performance
1. t2.micro	1 vCPU	1 GiB	Low to Moderate

2.12. Нажмите "Create Auto Scaling group"

Step 5: Add notifications Edit

Notifications

No notifications

Step 6: Add tags Edit

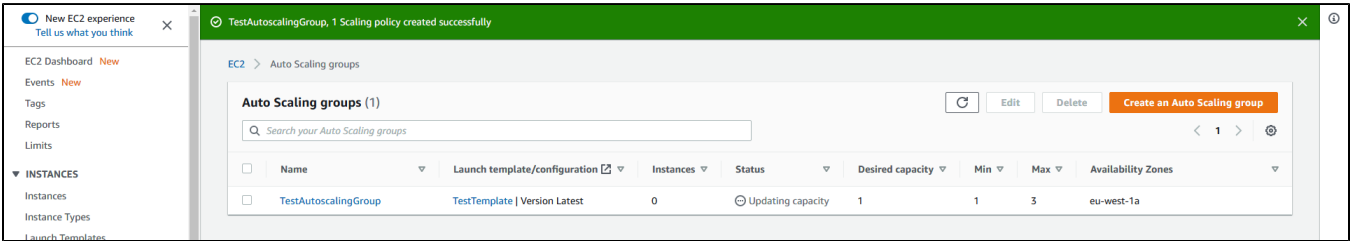
Tags (0)

Key	Value	Tag new instances
No tags		

Cancel

Create Auto Scaling group

Будет создана группа масштабирования, и запущен один экземпляр в группе



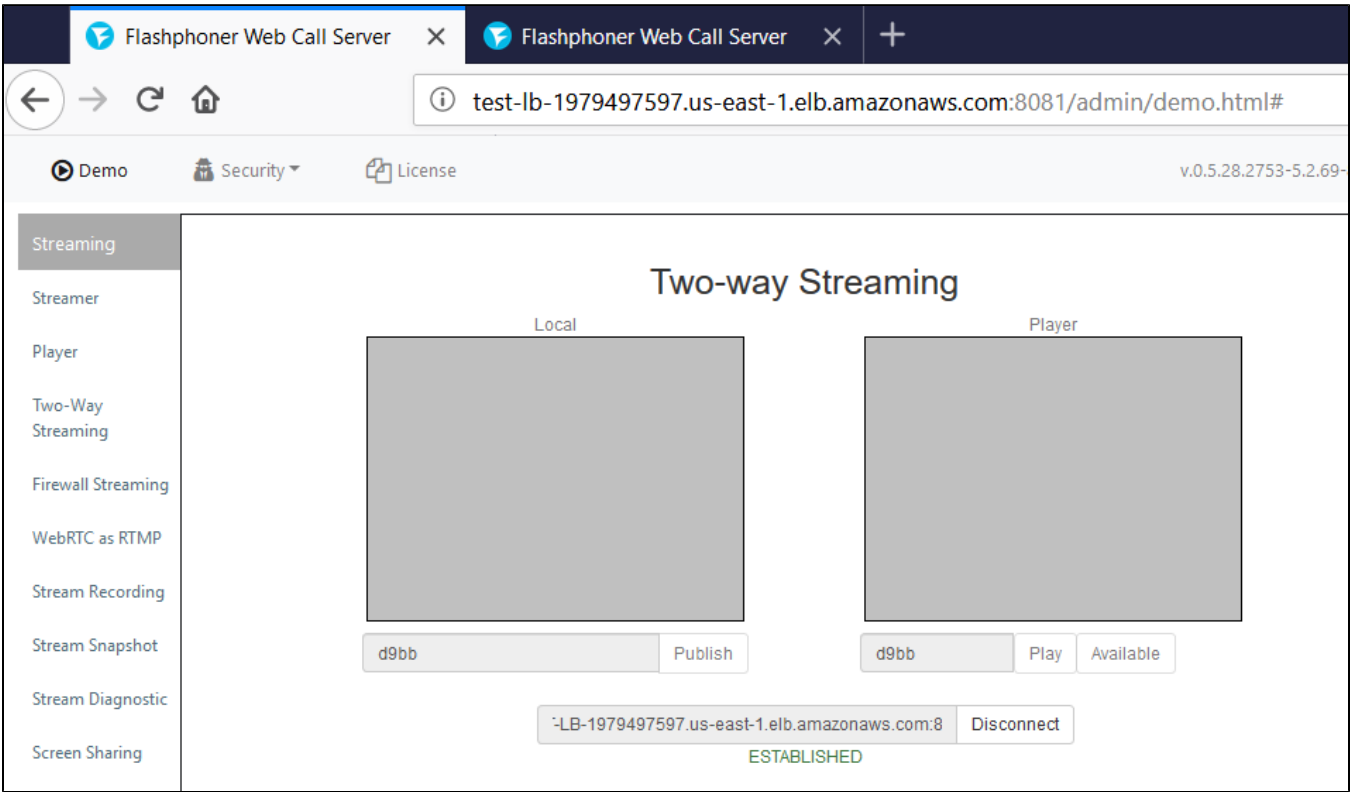
2.13. Измените параметры порта созданного балансировщика нагрузки, как описано [выше](#)

Тестирование

Если в балансировщике еще не запущено ни одного экземпляра сервера, новый экземпляр запустится в момент, когда будет создана группа масштабирования, получающая трафик от этого балансировщика. При выполнении условий масштабирования, будут созданы следующие экземпляры (Для загрузки процессора в ходе тестирования, может использоваться публикация потока с транскодингом, например, публикация RTMP потока в [автоматически создаваемый микшер](#)). Все запущенные экземпляры добавляются в соответствующий балансировщик.

Если экземпляр сервера (один или более из добавленных в балансировщик) работает, с ним может быть установлено WebSocket соединение, например `ws://<Load balancer DNS name>:8080`.

Пример [Two-way Streaming](#), (открытый по адресу балансировщика или конкретного экземпляра) может быть использован для установки WebSocket соединения:



Чтобы убедиться, что соединения распределяются между активными экземплярами в балансировщике нагрузки, используйте страницу статистики: `http://WCS_ADDRESS:8081/?action=stat`

Откройте страницу каждого экземпляра для контроля значения `connection_websocket`:

