

# Load balancing using Amazon Web Services – Auto Scaling feature

## 1. Create launch configuration

The screenshot shows the 'Create Launch Configuration' wizard in the AWS Management Console, specifically the '1. Choose AMI' step. The top navigation bar includes 'Services', 'Resource Groups', and user information. The wizard progress bar shows six steps: 1. Choose AMI, 2. Choose Instance Type, 3. Configure details, 4. Add Storage, 5. Configure Security Group, and 6. Review. The main heading is 'Create Launch Configuration' with a 'Cancel and Exit' link. Below the heading, a description states: 'An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.' The 'Quick Start' section on the left has a 'My AMIs' tab selected. The main area displays three AMIs: 'Amazon Linux AMI 2016.09.1 (HVM), SSD Volume Type - ami-9fc7b0f0', 'Red Hat Enterprise Linux 7.2 (HVM), SSD Volume Type - ami-cdbdd7a2', and 'SUSE Linux Enterprise Server 12 SP2 (HVM), SSD Volume Type - ami-8f8afde0'. Each AMI entry includes a description, root device type, virtualization type, and a 'Select' button.

The screenshot shows the 'Create Launch Configuration' wizard in the AWS Management Console, specifically the '3. Configure details' step. The top navigation bar and wizard progress bar are consistent with the previous step. The main heading is 'Create Launch Configuration'. The 'Name' field is set to 'MyLaunchConfig-01'. The 'Purchasing option' section has 'Request Spot Instances' unchecked. The 'IAM role' is set to 'None'. The 'Monitoring' section has 'Enable CloudWatch detailed monitoring' unchecked, with a 'Learn more' link. Below this, a description of a security group is provided: 'A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.' The 'Assign a security group' section has 'Create a new security group' selected. The 'Security group name' is 'AutoScaling-Security-Group-4' and the 'Description' is 'AutoScaling-Security-Group-4 (2015-11-26 08:19:19.290-06:00)'. At the bottom, a table lists the security group rules. The first rule is for 'SSH' (Type), 'TCP' (Protocol), 'Port Range' 22, and 'Source' 'Anywhere' (0.0.0.0/0). An 'Add Rule' button is located below the table.

| Type | Protocol | Port Range | Source             |
|------|----------|------------|--------------------|
| SSH  | TCP      | 22         | Anywhere 0.0.0.0/0 |

## 2. Create auto scaling group

## Create Auto Scaling Group

**Launch Configuration** ⓘ MyLaunchConfig-01

**Group name** ⓘ MyAutoScaleGroup-01

---

**Group size** ⓘ Start with  instances

---

**Network** ⓘ vpc-ca68c5ae (172.31.0.0/16) | Default (default) [Create new VPC](#)

**Subnet** ⓘ

subnet-61838638(172.31.16.0/20) | Default in us-east-1d x

subnet-cfda03f2(172.31.32.0/20) | Default in us-east-1e x

subnet-51a44627(172.31.0.0/20) | Default in us-east-1c x

subnet-771a2f5c(172.31.48.0/20) | Default in us-east-1b x

[Create new subnet](#)

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

## Create Auto Scaling Group

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

### Increase Group Size

**Name:**

**Execute policy when:** awsec2-MyAutoScaleGroup-01-High-CPU-Utilization [Edit](#) [Remove](#)  
breaches the alarm threshold: CPUUtilization <= 20 for 60 seconds  
for the metric dimensions AutoScalingGroupName = MyAutoScaleGroup-01

**Take the action:**

[Add step](#) ⓘ

**Instances need:**  seconds to warm up after each step

[Create a simple scaling policy](#) ⓘ

### Decrease Group Size

**Name:**

**Execute policy when:** awsec2-MyAutoScaleGroup-01-High-CPU-Utilization [Edit](#) [Remove](#)  
breaches the alarm threshold: CPUUtilization <= 20 for 60 seconds  
for the metric dimensions AutoScalingGroupName = MyAutoScaleGroup-01

**Take the action:**

[Add step](#) ⓘ

## Create Auto Scaling Group

Configure your Auto Scaling group to send notifications to a specified endpoint, such as an email address, whenever a specified event takes place, including: successful launch of an instance, failed instance launch, instance termination, and failed instance termination.

If you created a new topic, check your email for a confirmation message and click the included link to confirm your subscription. Notifications can only be sent to confirmed addresses.

Send a notification to:  [use existing topic](#) 

With these recipients:

Whenever instances:

- ☒ launch
- ☒ terminate
- ☒ fail to launch
- ☒ fail to terminate

[Add notification](#)

## Create Auto Scaling Group

Please review your Auto Scaling group details. You can go back to edit changes for each section. Click **Create Auto Scaling group** to complete the creation of an Auto Scaling group.

### Auto Scaling Group Details

[Edit details](#)

|                           |   |
|---------------------------|---|
| Group name                | MyAutoScaleGroup-01   |
| Group size                | 1   |
| Minimum Group Size        | 1   |
| Maximum Group Size        | 4   |
| Subnet(s)                 | subnet-61838638,subnet-cfda03f2,subnet-51a44627,subnet-771a2f5c |
| Health Check Grace Period | 300   |
| Detailed Monitoring       | No  |

### Scaling Policies

[Edit scaling policies](#)

|                     |   |
|---------------------|---|
| Increase Group Size | With alarm = awssec2-MyAutoScaleGroup-01-High-CPU-Utilization; Add 1 instances and 300 seconds for instances to warm up |
| Decrease Group Size | With alarm = awssec2-MyAutoScaleGroup-01-High-CPU-Utilization; Remove 1 instances                                       |

## Auto Scaling group creation status

✓ **Successfully created Auto Scaling group**  
[View creation log](#)








### View

[View your Auto Scaling groups](#)  
[View your launch configurations](#)

► Here are some helpful resources to get you started

[Close](#)

## 3. Monitor the instances that are being created

| <input type="checkbox"/>            | Name                 | Instance ID | Instance Type | Availability Zone | Instance State | Status Checks     | Alarm Status | Public DNS  |
|-------------------------------------|----------------------|-------------|---------------|-------------------|----------------|-------------------|--------------|---|
| <input type="checkbox"/>            |                      | i-22f11894  | t2.micro      | us-east-1c        | terminated     |                   | None         |                |
| <input type="checkbox"/>            |                      | i-89aeee3e  | t2.micro      | us-east-1c        | stopped        |                   | None         |                |
| <input type="checkbox"/>            | Autoscale-webservers | i-e393061a  | t2.micro      | us-east-1e        | running        | Initializing      | None         |  ec2-54-152-1f |
| <input type="checkbox"/>            | Autoscale-webservers | i-bfd23b09  | t2.micro      | us-east-1c        | running        | Initializing      | None         |  ec2-54-86-73- |
| <input checked="" type="checkbox"/> |                      | i-33016a83  | t2.micro      | us-east-1d        | terminated     |                   | None         |                |
| <input type="checkbox"/>            | Webserver            | i-b3b91c0d  | t2.micro      | us-east-1b        | running        | 2/2 checks passed | None         |  ec2-54-173-41 |
| <input type="checkbox"/>            | Autoscale-webservers | i-446e05f4  | t2.micro      | us-east-1d        | running        | 2/2 checks passed | None         |  ec2-52-91-24' |

Instance i-446e05f4 is created.

## 4. Increase load at the instance using a script.

```
Downloads — root@ip-172-31-18-86:/home/ec2-user — ssh — 117x27
 7 root      20    0    0    0    0 S  0.0  0.0  0:00.02 rcu_sched
 8 root      20    0    0    0    0 S  0.0  0.0  0:00.00 rcu_bh
 9 root      RT    0    0    0    0 S  0.0  0.0  0:00.00 migration/0
10 root      0 -20    0    0    0 S  0.0  0.0  0:00.00 khelper
11 root      20    0    0    0    0 S  0.0  0.0  0:00.00 kdevtmpfs
12 root      0 -20    0    0    0 S  0.0  0.0  0:00.00 netns
13 root      0 -20    0    0    0 S  0.0  0.0  0:00.00 perf
14 root      20    0    0    0    0 S  0.0  0.0  0:00.00 kworker/u30:1
16 root      20    0    0    0    0 S  0.0  0.0  0:00.01 xenwatch
21 root      20    0    0    0    0 S  0.0  0.0  0:00.00 xenbus
22 root      20    0    0    0    0 S  0.0  0.0  0:00.03 kworker/0:1
124 root     20    0    0    0    0 S  0.0  0.0  0:00.00 khungtaskd
125 root      0 -20    0    0    0 S  0.0  0.0  0:00.00 writeback
127 root     25    5    0    0    0 S  0.0  0.0  0:00.00 ksmd

[root@ip-172-31-18-86 ec2-user]# ls -ltr
total 4
-rw-r--r-- 1 root root 3621 Nov 26 14:30 dd.sh
[root@ip-172-31-18-86 ec2-user]# chmod 755 dd.sh
[root@ip-172-31-18-86 ec2-user]# nohup ./dd.sh &
[1] 2491
[root@ip-172-31-18-86 ec2-user]# nohup: ignoring input and appending output to 'nohup.out'

[root@ip-172-31-18-86 ec2-user]# nohup ./dd.sh &
[2] 2499
[root@ip-172-31-18-86 ec2-user]# nohup: ignoring input and appending output to 'nohup.out'

[root@ip-172-31-18-86 ec2-user]# nohup ./dd.sh &
```

CPU usage has increased

```
top - 14:31:03 up 2 min, 1 user, load average: 0.75, 0.23, 0.09
Tasks: 90 total, 5 running, 85 sleeping, 0 stopped, 0 zombie
Cpu(s): 3.1%us, 89.7%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%si, 7.2%st
Mem: 1019452k total, 159272k used, 860180k free, 8356k buffers
Swap: 0k total, 0k used, 0k free, 92864k cached


  PID USER      PR  NI  VIRT  RES  SHR  S  %CPU  %MEM    TIME+  COMMAND
 2547 root      20   0 105m  756  684  R   7.7   0.1   0:00.23 dd
 2548 root      20   0 105m  744  668  R   5.0   0.1   0:00.15 dd
 2549 root      20   0 105m  708  632  R   5.0   0.1   0:00.15 dd
 2550 root      20   0 105m  660  588  R   4.7   0.1   0:00.14 dd
    1 root      20   0 19612 2468 2148  S   0.0   0.2   0:00.83 init
    2 root      20   0    0    0    0  S   0.0   0.0   0:00.00 kthreadd
    3 root      20   0    0    0    0  S   0.0   0.0   0:00.00 ksoftirqd/0
    4 root      20   0    0    0    0  S   0.0   0.0   0:00.00 kworker/0:0
    5 root      0 -20    0    0    0  S   0.0   0.0   0:00.00 kworker/0:0H
    6 root      20   0    0    0    0  S   0.0   0.0   0:00.00 kworker/u30:0
    7 root      20   0    0    0    0  S   0.0   0.0   0:00.02 rcu_sched
    8 root      20   0    0    0    0  S   0.0   0.0   0:00.00 rcu_bh
    9 root      RT    0    0    0    0  S   0.0   0.0   0:00.00 migration/0
   10 root      0 -20    0    0    0  S   0.0   0.0   0:00.00 khelper
   11 root      20   0    0    0    0  S   0.0   0.0   0:00.00 kdevtmpfs
   12 root      0 -20    0    0    0  S   0.0   0.0   0:00.00 netns
   13 root      0 -20    0    0    0  S   0.0   0.0   0:00.00 perf
   14 root      20   0    0    0    0  S   0.0   0.0   0:00.00 kworker/u30:1
   16 root      20   0    0    0    0  S   0.0   0.0   0:00.01 xenwatch
   21 root      20   0    0    0    0  S   0.0   0.0   0:00.00 xenbus
```

## 5. New instance is created

| <input type="checkbox"/>            | Name                 | Instance ID | Instance Type | Availability Zone | Instance State | Status Checks     | Alarm Status | Public DNS    |
|-------------------------------------|----------------------|-------------|---------------|-------------------|----------------|-------------------|--------------|---------------|
| <input type="checkbox"/>            |                      | i-22f11894  | t2.micro      | us-east-1c        | terminated     |                   | None         |               |
| <input type="checkbox"/>            |                      | i-89aeee3e  | t2.micro      | us-east-1c        | stopped        |                   | None         |               |
| <input checked="" type="checkbox"/> | Autoscale-webservers | i-bfd23b09  | t2.micro      | us-east-1c        | running        | Initializing      | None         | ec2-54-86-73- |
| <input type="checkbox"/>            |                      | i-33016a83  | t2.micro      | us-east-1d        | terminated     |                   | None         |               |
| <input type="checkbox"/>            | Webserver            | i-b3b91c0d  | t2.micro      | us-east-1b        | running        | 2/2 checks passed | None         | ec2-54-173-4: |
| <input type="checkbox"/>            | Autoscale-webservers | i-446e05f4  | t2.micro      | us-east-1d        | running        | 2/2 checks passed | None         | ec2-52-91-24' |

Notification is received on the email ID.





**AWS Notifications** <no-reply@sns.amazonaws.com>  
to me

Apr 3 (13 days ago)

Service: AWS Auto Scaling  
Time: 2018-04-03T10:39:17.442Z  
RequestId: 15058a2c-59b0-aa7a-c2c8-06441676c9e4  
Event: autoscaling\_EC2\_INSTANCE\_LAUNCH  
AccountID: 413149532607  
AutoScalingGroupName: myautoscale1  
AutoScalingGroupARN: arn:aws:autoscaling:us-east-1:413149532607:autoScalingGroup:e4242dbf-24cf-4022-a7c1-aa02578f1e9a:autoScalingGroupName/myautoscale1  
ActivityId: 15058a2c-59b0-aa7a-c2c8-06441676c9e4  
Description: Launching a new EC2 instance: i-078b777cfe785a6a3  
Cause: At 2018-04-03T10:38:43Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 2 to 3.  
StartTime: 2018-04-03T10:38:45.290Z  
EndTime: 2018-04-03T10:39:17.442Z  
StatusCode: InProgress  
StatusMessage:  
Progress: 50  
EC2InstanceID: i-078b777cfe785a6a3  
Details: {"Subnet ID": "subnet-57a4bf1c", "Availability Zone": "us-east-1a"}

4 instances are created.

| <input type="checkbox"/>            | Name                | Instance ID | Instance Type | Availability Zone | Instance State | Status Checks     | Alarm Status | Public DNS     |
|-------------------------------------|---------------------|-------------|---------------|-------------------|----------------|-------------------|--------------|----------------|
| <input type="checkbox"/>            |                     | i-22f11894  | t2.micro      | us-east-1c        | terminated     |                   | None         |                |
| <input type="checkbox"/>            |                     | i-89aeee3e  | t2.micro      | us-east-1c        | stopped        |                   | None         |                |
| <input type="checkbox"/>            | Autoscale-webserver | i-a393061a  | t2.micro      | us-east-1e        | running        | Initializing      | None         | ec2-54-152-11f |
| <input type="checkbox"/>            | Autoscale-webserver | i-bfd23b09  | t2.micro      | us-east-1c        | running        | Initializing      | None         | ec2-54-86-73-  |
| <input type="checkbox"/>            |                     | i-33016a83  | t2.micro      | us-east-1d        | terminated     |                   | None         |                |
| <input type="checkbox"/>            | Webserver           | i-b3b91c0d  | t2.micro      | us-east-1b        | running        | 2/2 checks passed | None         | ec2-54-173-45  |
| <input checked="" type="checkbox"/> | Autoscale-webserver | i-446e0514  | t2.micro      | us-east-1d        | running        | 2/2 checks passed | None         | ec2-52-91-24'  |

CPU usage goes low after sometime

Downloads — root@ip-172-31-18-86:/home/ec2-user — ssh — 117x27

Tasks: 80 total, 1 running, 79 sleeping, 0 stopped, 0 zombie  
Cpu(s): 0.0%us, 0.0%sy, 0.0%ni, 100.0%id, 0.0%wa, 0.0%hi, 0.0%si, 0.0%st  
Mem: 1019452k total, 158936k used, 860516k free, 9592k buffers  
Swap: 0k total, 0k used, 0k free, 93096k cached

| PID | USER | PR | NI  | VIRT  | RES  | SHR  | S | %CPU | %MEM | TIME+   | COMMAND       |
|-----|------|----|-----|-------|------|------|---|------|------|---------|---------------|
| 1   | root | 20 | 0   | 19612 | 2468 | 2148 | S | 0.0  | 0.2  | 0:00.03 | init          |
| 2   | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | kthreadd      |
| 3   | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.02 | ksoftirqd/0   |
| 5   | root | 0  | -20 | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | kworker/0:0H  |
| 6   | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | kworker/u30:0 |
| 7   | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | rcu_sched     |
| 8   | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | rcu_bh        |
| 9   | root | RT | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | migration/0   |
| 10  | root | 0  | -20 | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | khelper       |
| 11  | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | kdevtmpfs     |
| 12  | root | 0  | -20 | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | netns         |
| 13  | root | 0  | -20 | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | perf          |
| 14  | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | kworker/u30:1 |
| 16  | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.01 | xenwatch      |
| 21  | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | xenbus        |
| 22  | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.06 | kworker/0:1   |
| 124 | root | 20 | 0   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | khungtaskd    |
| 125 | root | 0  | -20 | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | writeback     |
| 127 | root | 25 | 5   | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | ksmd          |
| 128 | root | 39 | 19  | 0     | 0    | 0    | S | 0.0  | 0.0  | 0:00.00 | khugepaged    |

[root@ip-172-31-18-86 ec2-user]#

Instance starts terminating as CPU usage is low.

| <input type="checkbox"/>            | Name                | Instance ID | Instance Type | Availability Zone | Instance State | Status Checks     | Alarm Status | Public DNS     |
|-------------------------------------|---------------------|-------------|---------------|-------------------|----------------|-------------------|--------------|----------------|
| <input type="checkbox"/>            |                     | i-22f11894  | t2.micro      | us-east-1c        | terminated     |                   | None         |                |
| <input type="checkbox"/>            |                     | i-89aeee3e  | t2.micro      | us-east-1c        | stopped        |                   | None         |                |
| <input type="checkbox"/>            | Autoscale-webserver | i-a393061a  | t2.micro      | us-east-1e        | running        | Initializing      | None         | ec2-54-152-11f |
| <input type="checkbox"/>            | Autoscale-webserver | i-bfd23b09  | t2.micro      | us-east-1c        | running        | Initializing      | None         | ec2-54-86-73-  |
| <input checked="" type="checkbox"/> |                     | i-33016a83  | t2.micro      | us-east-1d        | terminated     |                   | None         |                |
| <input type="checkbox"/>            | Webserver           | i-b3b91c0d  | t2.micro      | us-east-1b        | running        | 2/2 checks passed | None         | ec2-54-173-45  |
| <input type="checkbox"/>            | Autoscale-webserver | i-446e0514  | t2.micro      | us-east-1d        | running        | 2/2 checks passed | None         | ec2-52-91-24'  |