

# Exercise 3

## *Configuring a Load Balancer Stress Testing*

### **Prior Knowledge**

Unix Command Line Shell

Exercise 2: Auto Scaling groups and Launch Configurations

### **Learning Objectives**

Creating an elastically scaled system in the cloud

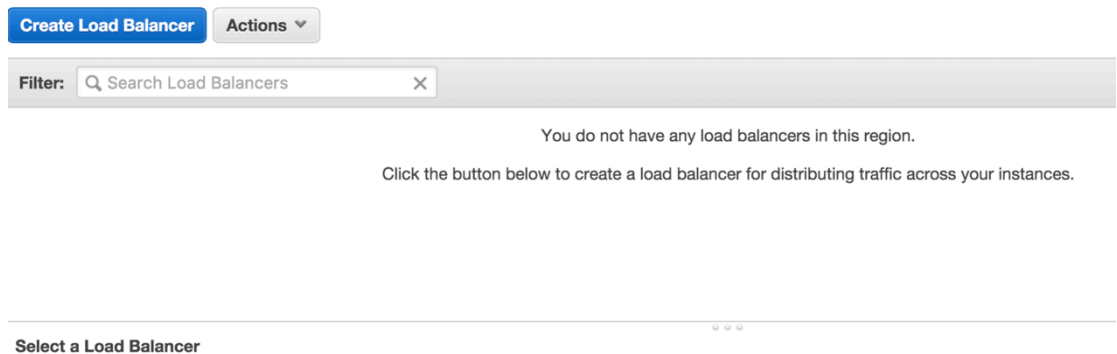
How to stress test using Linux siege command

### **Software Requirements**

Browser and AWS account, previous configuration from Exercise 2

### **Part A: Setting up a Load Balancer and ELB Auto Scale Group**

1. Go to the AWS Console and then the EC2 Console.
2. Near the bottom of the left hand menu, find Load Balancers and Click on it. You will see something like this (although other students may have created load balancers that will show up).



3. Click **Create Load Balancer**
4. Choose **Classic Load Balancer**

5. In the screen following:
  - a. Set name to *userid-elb* (e.g. *oxclo02-elb*)
  - b. Leave the Load Balancer protocol as HTTP, etc, except change the **Instance Port** to 8080.This will mean that traffic coming to the LB will be sent to port 8080 on the instance servers.

Load Balancer Protocol	Load Balancer Port	Instance Protocol	Instance Port
HTTP	80	HTTP	80

6. Click **Next: Assign Security Groups**
7. Select **Create a New Security Group**
8. Give it the name *userid-elb-sg* (e.g. *oxclo02-elb-sg*)
9. Make sure the rule says:  
HTTP TCP 80 Anywhere 0.0.0.0/0

Type	Protocol	Port Range	Source
HTTP	TCP	80	Anywhere 0.0.0.0/0

10. Click **Next: Configure Security Settings**
11. Ignore the warning and click: **Next: Configure Health Check**

12. Change the settings as follows:

- a. Ping Protocol: HTTP
- b. Ping Port: 8080
- c. Ping Path: /
- d. Response Timeout: 5
- e. Interval: 10
- f. Unhealthy threshold: 5
- g. Healthy threshold: 5

The screenshot shows the configuration interface for an AWS Auto Scaling Group. At the top, there are three fields: 'Ping Protocol' set to 'HTTP', 'Ping Port' set to '8080', and 'Ping Path' set to '/'. Below these is a section titled 'Advanced Details' which is expanded. It contains four rows of settings, each with an information icon, a label, a value field, and a unit: 'Response Timeout' with value '5' and unit 'seconds', 'Health Check Interval' with value '10' and unit 'seconds', 'Unhealthy Threshold' with value '5', and 'Healthy Threshold' with value '5'.

13. Click **Next: Add EC2 Instances**

14. Do NOT add any instances! Click **Next: Add Tags**

15. Add the tag with Key/Value: Name / *userid-asi*

16. Click **Review and Create** then **Create**

17. Click **Close**

18. Now let's create our AutoScaling Group

19. Go back to creating an Auto Scale Group like last time. (**Auto Scaling Groups -> Create Auto Scaling Group**)

20. Create from an existing Launch Configuration and choose your own launch config that you previously created. Click **Next Step**

21. On the following screen:

- a. Give it a group name of *userid-asg* (e.g. *oxclo02-asg*)
- b. Add one or more subnets as before
- c. Expand the **Advanced Details**
- d. Click **Receive Traffic from Elastic Load Balancers**
- e. Select your own Load Balancer from the options

- f. Change the Health Check type to ELB
- g. Leave the Grace period as 300 seconds
- h. Click **Next: Configure Scaling Policies**

22. On the following screen

- a. Select **Use scaling policies....**
- b. Change it to support scaling between 1 and 4 instances
- c. Click Add New Alarm
- d. If you want notifications, choose your own topic that you defined before.
- e. Change the Alarm to fire when the CPU utilization is  $\geq 35\%$  for more than 1 minute (we want to see scaling, so this is deliberately low). Note that the picture below shows 40%, but I recommend using 35% to ensure that we cause enough work to see scaling.

**Create Alarm**

You can use CloudWatch alarms to be notified automatically whenever metric data reaches a level you define.  
To edit an alarm, first choose whom to notify and then define when the notification should be sent.

☒ **Send a notification to:**  [cancel](#)

**With these recipients:**

**Whenever:** Average of CPU Utilization

**Is:**  $\geq$  40 Percent

**For at least:** 1 consecutive period(s) of 1 Minute

**Name of alarm:**

**CPU Utilization Percent**

40  
30  
20  
10  
0

1/1 00:00 1/1 00:00 1/1 00:00

☒ blah

[Cancel](#) [Create Alarm](#)

- f. Click **Create Alarm**

23. Now update the rule to **Add 1 instance**

24. Set **Instances need 300 seconds to warm up after each step**

25. Create a similar Alarm for when CPU utilization is  $\leq 25\%$  for 2 minutes, and change the rule to Remove 1 instance.

26. It should look like (again, note that the picture shows different %ages)

The screenshot shows the AWS Auto Scaling console with two scaling policies configured for an Auto Scaling Group.

**Increase Group Size Policy:**

- Name:** Increase Group Size
- Execute policy when:** awsec2-oxsoa02-CPU-Utilization [Edit](#) [Remove](#)  
breaches the alarm threshold: CPUUtilization >= 40 for 60 seconds  
for the metric dimensions AutoScalingGroupName = blah
- Take the action:** Add 1 instances when 40 <= CPUUtilization < +infinity  
[Add step](#) ⓘ
- Instances need:** 300 seconds to warm up after each step
- [Create a simple scaling policy](#) ⓘ

**Decrease Group Size Policy:**

- Name:** Decrease Group Size
- Execute policy when:** awsec2-oxclo03-asg-CPU-Utilization [Add new alarm](#)  
breaches the alarm threshold: CPUUtilization <= 30 for 2 consecutive periods of 60 seconds  
for the metric dimensions AutoScalingGroupName = oxclo03-asg
- Take the action:** Remove 1 instances when 30 >= CPUUtilization > -infinity  
[Add step](#) ⓘ

27. Click **Next: Configure Notifications**

28. Click **Next: Configure Tags**

29. Add the tag: Name / *userid*-asi

30. Click **Review**

31. Click **Create Autoscaling Group**

32. Go and see if your instances are being started.

## PART B – Stress testing

33. Navigate to view your ELB's dashboard page. You can find the DNS address of your ELB this way:

The screenshot shows the AWS Elastic Load Balancing console for a load balancer named 'oxclo02-elb'.

**Load balancer:** oxclo02-elb

**Description** | Instances | Health Check | Monitoring | Security | Listeners | Tags

**DNS Name:** oxclo02-elb-137471382.eu-west-1.elb.amazonaws.com (A Record)

34. After the system has warmed up and your instance is running, it will eventually be tested by the ELB and become **In-Service**. You should see something like this:

Load balancer: **oxclo02-elb**

Description

**Instances**

Health Check

Monitoring

Security

Listeners

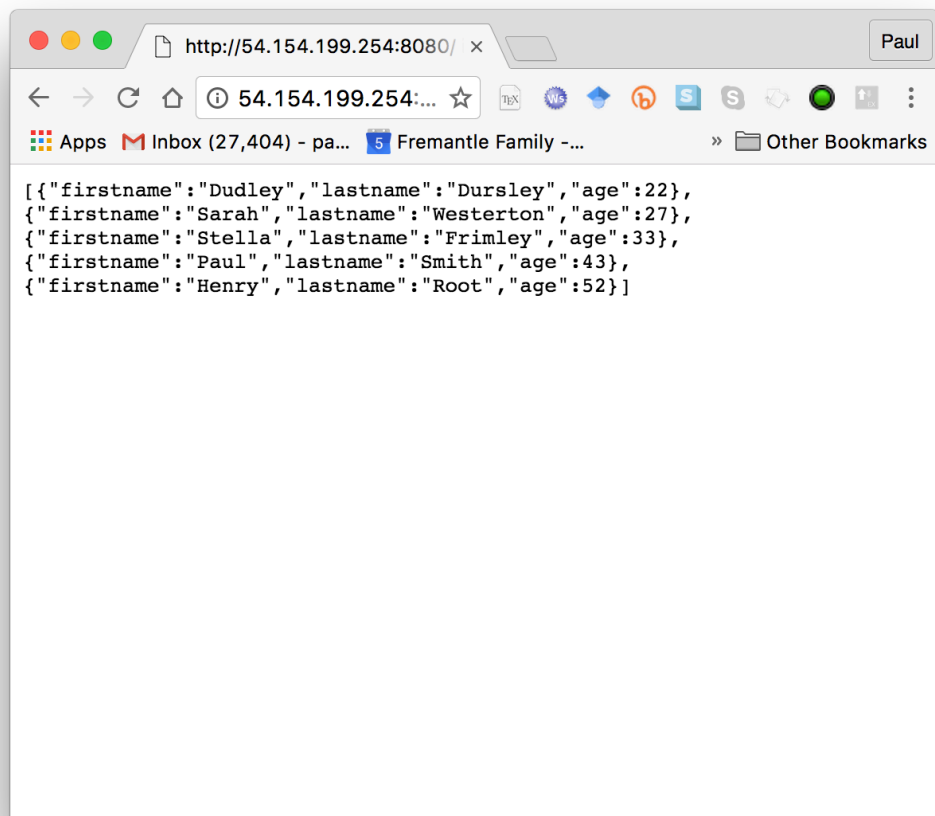
Tags

Connection Draining: Enabled, 300 seconds ([Edit](#))

[Edit Instances](#)

Instance ID	Name	Availability Zone	Status
<a href="#">i-3483498d</a>	oxclo02-asg	eu-west-1a	InService <a href="#">i</a>

35. Once you have an InService instance, copy and paste the DNS name into the address bar of your browser. You should see JSON returned from the node.js app.



Notice this is now available on port 80 and no longer using 8080, because the load balancer listens on 80.

36. We are going to create a new instance in the same subnet to stress test the servers from. We could do it from here, but we will take out network delays if we can do it within the Amazon EC2 network.

37. Using the EC2 Launch wizard like before, start a new instance with the following settings:
- a. **Ubuntu Server 16.04 LTS (HVM)**
  - b. **t2.medium** (we want a beefier machine to be able to drive our nodes hard)
  - c. User Data: please cut and paste from <http://freo.me/oxclo-siege-ud>
  - d. This simply installs the latest version of *siege* and sets correct parameters for the OS to handle this. (the version in the Ubuntu repo is out of date and buggy unfortunately). Later we will see how Docker would solve this problem better....
  - e. Tag Name: *userid-siege*
  - f. Security Group: node-security-group
  - g. Your existing SSH Key

38. Check the instance is running in the EC2 dashboard and then SSH into the instance as in Exercise 1.

39. Accept the fingerprint as before.

40. In the SSH session type:

```
siege -c100 -t15m http://your-lb-dns-goes-here
```

e.g

```
siege -c100 -t15m http://oxclo02-elb-137471382.eu-west-1.elb.amazonaws.com
```

41. You should see something like:

```
New configuration template added to /home/ubuntu/.siege
Run siege -C to view the current settings in that file
[alert] Zip encoding disabled; siege requires zlib support to
enable it: No such file or directory
** SIEGE 4.0.3rc4
** Preparing 100 concurrent users for battle.
The server is now under siege...
```

42. This is basically hitting your Load Balancer with 100 concurrent clients for 15 minutes. This should be long enough to see the behaviour we want.

43. You may also see messages like:

```
[error] A temporary resolution error for oxclo01-elb-2104065837.eu-
west-1.elb.amazonaws.com
You can ignore these.
```

44. Unless we run out of network bandwidth, this should push the instances' average CPU above 40% and cause the Scaling Group to start another server.

45. Assuming all is well you should see a new instance spawned shortly.

46. You can also check the Auto Scaling Group's Activity History

Auto Scaling Group: oxclo03-asg

Details Activity History Scaling Policies Instances Notifications Tags

Filter: Any Status  1 to 2 of 2 History Items

Status	Description	Start Time	End Time
Waiting for instance warmup	Launching a new EC2 instance: i-fbd8ef42	2015 November 17 14:16:55 UTC	
Successful	Launching a new EC2 instance: i-f2bf754b	2015 November 17 13:24:22 UTC	2015 November 17 13:25:26 UTC

47. And the Elastic Load Balancer's instances

Load balancer: oxclo02-elb

Description Instances Health Check Monitoring Security Listeners Tags

Connection Draining: Enabled, 300 seconds (Edit)

Edit Instances

Instance ID	Name	Availability Zone	Status	Actions
i-f2bf754b	oxclo02-asi	eu-west-1a	InService ⓘ	<a href="#">Remove from Load Balancer</a>
i-fbd8ef42	oxclo02-asi	eu-west-1b	OutOfService ⓘ	<a href="#">Remove from Load Balancer</a>



48. Once you have seen the scaling, you can end the siege if you like, by hitting Ctrl-C in the command line window:

```
Lifting the server siege...
Transactions:          613905 hits
Availability:          99.96 %
Elapsed time:          599.40 secs
Data transferred:     157.49 MB
Response time:         0.09 secs
Transaction rate:      1024.20 trans/sec
Throughput:            0.26 MB/sec
Concurrency:           95.61
Successful transactions: 613905
Failed transactions:    3
Longest transaction:    6.01
Shortest transaction:   0.00
```

49. Once the siege has ended, you should see the spare instance removed:

Auto Scaling Group: oxclo03-asg

Details

Activity History

Scaling Policies

Instances

Notifications

Tags

Filter: Any Status

Filter scaling history...

	Status	Description	Start Time
▶	Successful	Terminating EC2 instance: i-fbd8ef42	2015 November 17 14:24:24 UTC
▶	Successful	Launching a new EC2 instance: i-fbd8ef42	2015 November 17 14:16:55 UTC
▶	Successful	Launching a new EC2 instance: i-f2bf754b	2015 November 17 13:24:22 UTC

50. Once you have finished:

- Delete the autoscaling group
- Delete the load balancer
- Terminate the siege instance.
- Make sure that you have no further instances running in your name!

51. You have completed the exercise. Well done.

52. As an **extension**, come up with a plan to secure the cloud instances better through improved configuration of the security groups. Identify which systems need to talk to which, and then suggest a set of security groups that would allow this.