

Enhancing and Optimizing Computing using Elastic Load Balancing and Auto Scaling

1. Introduction

Cloud computing is a newly progressing technique which offers online computing resources, storage and permits users to organize applications with enhanced scalability, availability and fault tolerance. Cloud computing is about storing the stuff on remote servers instead of on own computers or other devices. This information can be retrieved using the internet with any device, everywhere in the world as long as that device can support cloud computing systems. The cloud computing system is comprised of a front-end, which is the client side and a back-end which is a collection of the servers and computers owned by a third party which stores the data. A central server which is a fragment of the back-end follows protocols and uses middleware to communicate between networked computers. Cloud computing accumulates all the computing resources and manages them automatically. Its characteristics describe a cloud computing system on-need self-service, pooling of resources and access to the internet. The elasticity of service availability and measurement of services utilized by individual users. Cloud computing is everywhere with tools like Google Drives replacing Microsoft Office, Amazon Web Services replacing traditional enterprise data storage, banking websites replacing branch offices and Dropbox storing all our data and files.

2. Load Balancing

Load balancing is a serious concern in cloud computing. With the increase in attractiveness of cloud computing among users, the load on the servers and the quantity of processing done is surging drastically. There are multiple nodes in the cloud, and due to the random allocation of a request made by the client to any node, the nodes become unevenly loaded. So to avoid the condition where some nodes are either severely loaded or under loaded, the load balancer will evenly divide the workload among all the nodes. Thus load balancing will equally distribute the workload among the nodes, and it can help in minimizing delays in communication, maximizing the throughput, minimizing execution time and maximizing resource utilization.

2.1 Advantages of Load Balancing

1. It should possess fault tolerance.
2. It should be capable of modifying itself according to any changes.
3. It should also maintain system stability.

2.2 Issues of Load Balancing

1. Load balancing becomes critical because, in the middle of execution, the processes may shift amongst nodes to ensure equal workload on the system.
2. For a load balancing scheme to be good it should be scalable, general and stable and should add minimal overhead to the system. These requirements are interdependent.

3. Auto Scaling

Auto Scaling is a method used in cloud computing, whereby the amount of computational resources in a server farm, typically

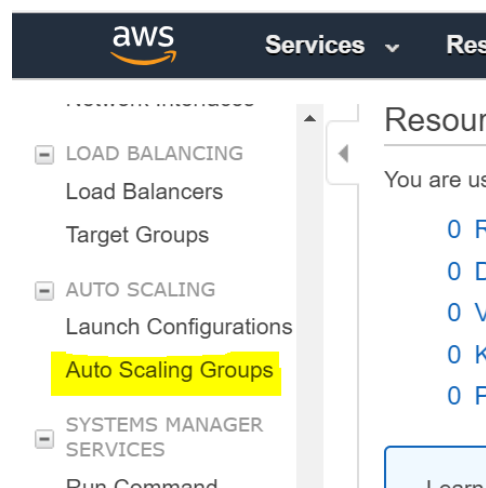
measured in terms of the number of active servers, scales automatically based on the load on the farm. It is closely related to, and builds upon, the idea of Load Balancing.

3.1 Advantages

1. For companies running their own web server infrastructure, auto scaling typically means allowing some servers to go to sleep during times of low load, saving on electricity costs (as well as water costs if water is being used to cool the machines).
2. For companies using infrastructure hosted in the cloud, auto scaling can mean lower bills, because most cloud providers charge based on total usage rather than maximum capacity.
3. Auto scaling can offer greater uptime and more availability in cases where production workloads are variable and unpredictable.
4. Auto scaling solutions, such as the one offered by Amazon Web Services, can also take care of replacing unhealthy instances and therefore protecting somewhat against hardware, network, and application failures.

4. Creating an Application Load Balancer and Auto Scaling

1. Open EC2 console.
2. Creating Auto Scaling group.
 - 2.1 Go to Auto Scaling Groups.



2.2 Create Auto Scaling group and Create launch Configuration.

2.3 Select Microsoft Windows Server 2012 R2 Base.

**Microsoft Windows Server 2012 R2 Base** - ami-e795bd88

Microsoft Windows 2012 R2 Standard edition with 64-bit architecture. [English]

Free tier eligible

Root device type: ebs Virtualization type: hvm

Select

64-bit

2.4 In create Launch Configuration select t2 small with 2GB RAM.

<input type="checkbox"/>	General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate
<input checked="" type="checkbox"/>	General purpose	t2.small	1	2	EBS only	-	Low to Moderate

2.5 Configure details (give your server name).

2.6 Add Storage to your Server.

2.7 In configure Security Group (give security group name and description) and add Rules as per your need.

Create Launch Configuration

Security group name:

Description:

Type ⓘ	Protocol ⓘ	Port Range ⓘ	Source ⓘ
RDP ▼	TCP	3389	Anywhere ▼ 0.0.0.
All traffic ▼	All	0 - 65535	Anywhere ▼ 0.0.0.
SSH ▼	TCP	22	Anywhere ▼ 0.0.0.
HTTPS ▼	TCP	443	Anywhere ▼ 0.0.0.
HTTP ▼	TCP	80	Anywhere ▼ 0.0.0.

Add Rule

2.8 Create Launch Configuration, download the new key pair and create launch.

2.9 Configure Auto Scaling group details.


1. Configure Auto Scaling group details 2. Configure scaling policies 3. Configure Notifications 4. Configure Tags 5. Review

Create Auto Scaling Group


Launch Configuration ⓘ serveronea

Group name ⓘ autoserveronea

Group size ⓘ Start with 1 instances

Network ⓘ vpc-1761237f (172.31.0.0/16) (default)  Create new VPC

Subnet ⓘ

subnet-503b9c1c(172.31.0.0/20) | Default in ap-south-1b
subnet-c6c3b5ae(172.31.16.0/20) | Default in ap-south-1a
|  Create new subnet

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

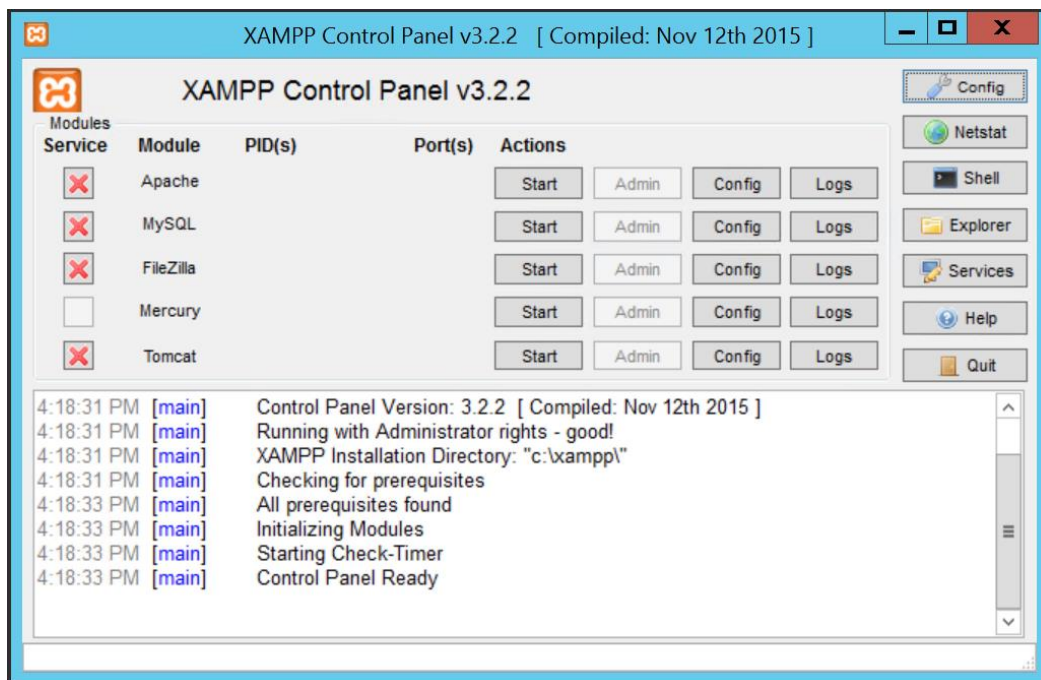
2.10 Now create auto scaling group and check whether the instance is running in EC2.

2.11 Now create another auto scaling group using the same process and name the instances.

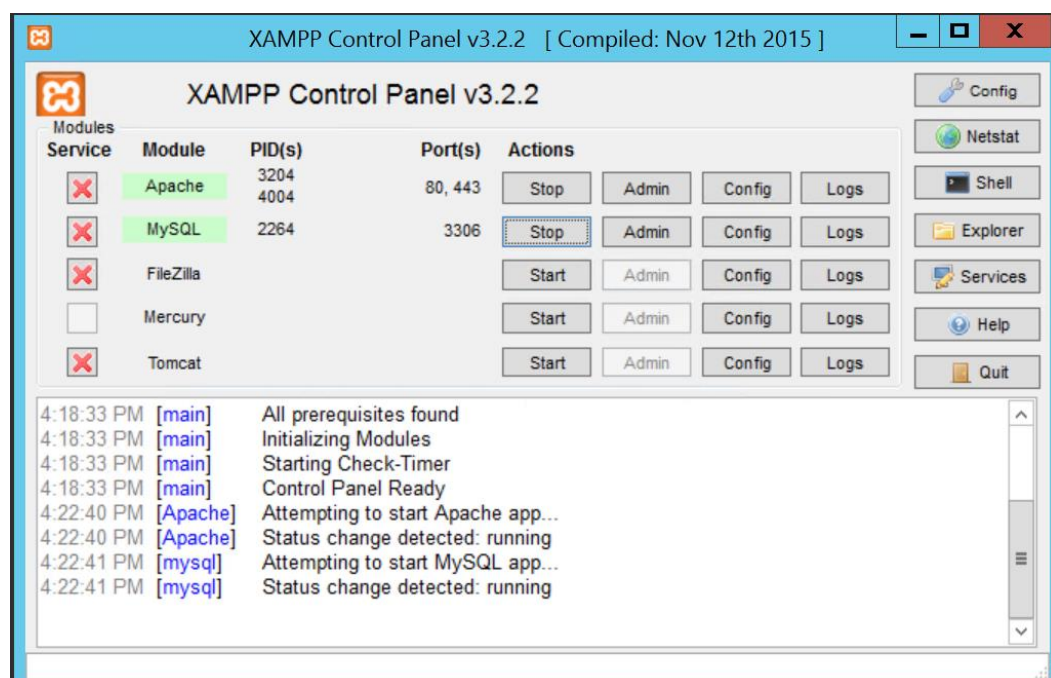
serveroneb	i-0e18a77826a3d879c	t2.small	ap-south-1b	running	Initializing	None
serveronea	i-0e18a71a4b91e9569	t2.small	ap-south-1b	running	2/2 checks...	None

2.12 Connect both the instances and disable the firewall.

2.13 Download xampp on one of the server.

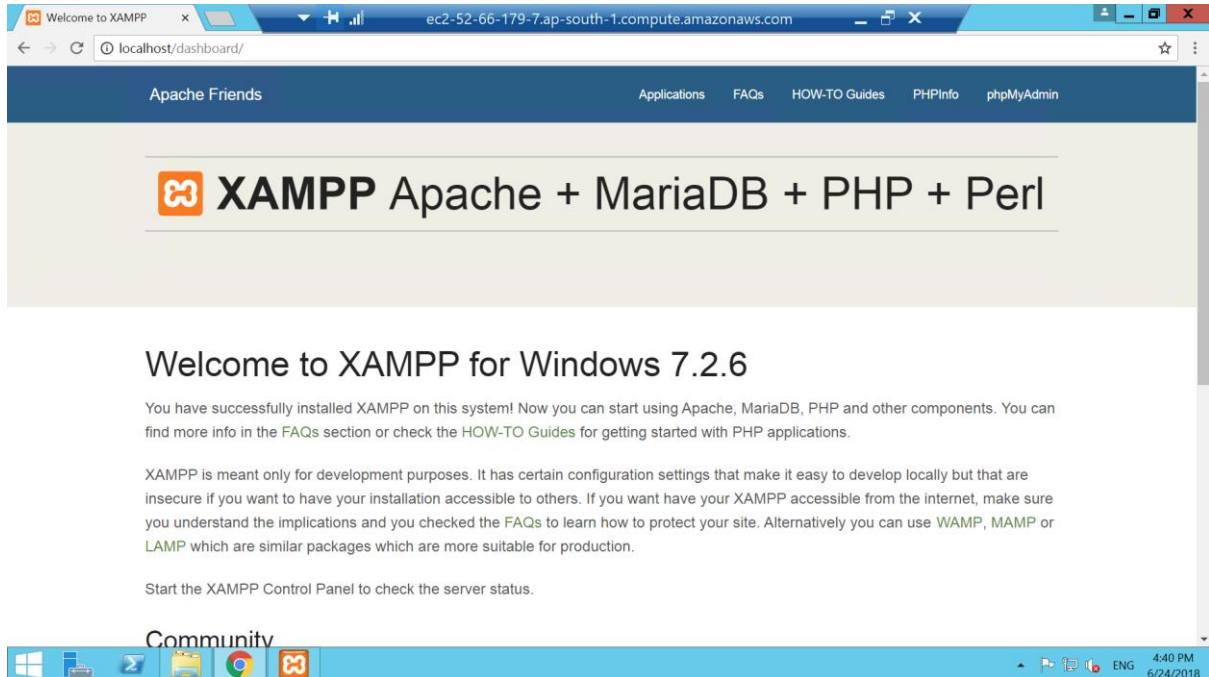


2.14 Start Apache and MySQL.

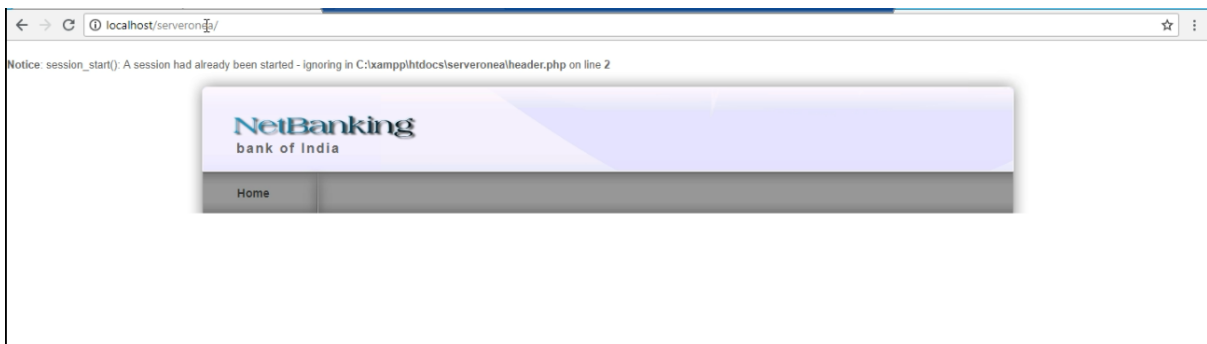


2.15 Copy your website to “C:\xampp\htdocs” by creating a new folder.

2.16 Go to the browser and type localhost



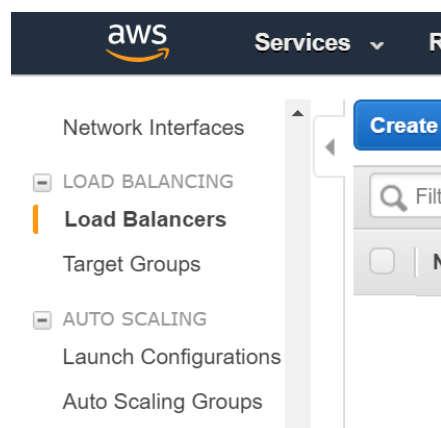
2.17 Now type localhost/'foldername'



2.18 Do the same thing to the second server and check.

Note: auto scaling is done now load balancer part

2.19 Go to Load Balancer



2.20 Create application load balancer and select all the availability zones

aws

Services

Resource Groups

★

1. Configure Load Balancer

2. Configure Security Settings

3. Configure Security Groups

4. Configure Routing

5. Register Targets

6. Review

Step 1: Configure Load Balancer

Name ⓘ

netbanking

Scheme ⓘ

internet-facing

internal

IP address type ⓘ

ipv4

Listeners

A listener is a process that checks for connection requests, using the protocol and port that you configured.

Load Balancer Protocol

Load Bal

HTTP

80

Add listener

Availability Zones

Specify the Availability Zones to enable for your load balancer. The load balancer routes traffic to the targets in these Availability Zones only. You can specify or

balancer.

VPC ⓘ

vpc-1761237f (172.31.0.0/16) (default)

<input type="checkbox"/>	Availability Zone	Subnet ID	Subnet IPv4 CIDR
<input checked="" type="checkbox"/>	ap-south-1a	subnet-c6c3b5ae	172.31.16.0/20
<input checked="" type="checkbox"/>	ap-south-1b	subnet-503b9c1c	172.31.0.0/20

2.21 Select security group.

<input checked="" type="checkbox"/>	sg-63083b08	default	default VPC security group
<input type="checkbox"/>	sg-06257f8593462b2e6	jenmas	jenmas
<input type="checkbox"/>	sg-0155b3dc7d081844e	load-balancer-wizard-1	load-balancer-wizard-1 created on 2018-06-01T12:42:28.798-07:00
<input type="checkbox"/>	sg-0f601bc66f457b037	scaleload	scaleload
<input checked="" type="checkbox"/>	sg-0e01bd7d2d96e580a	serveronea	serveronea

2.22 In register targets select both the servers and add to register

Instances

To register additional instances, select one or more running instances, specify a port, and then click Add. The default port is the port specified for the target group. If the instance is already registered on the specified port, you must specify a different port.

Add to registered

on port 80

Search Instances

X

<input type="checkbox"/>	Instance	Name	State	Security groups	Zone	Subnet ID	Subnet CIDR
<input checked="" type="checkbox"/>	i-0e18a77828a3d879c	serveroneb	running	serveronea	ap-south-1b	subnet-2e9f4c62	172.31.0.0/20
<input checked="" type="checkbox"/>	i-0e18a71a4b91e9569	serveronea	running	serveronea	ap-south-1b	subnet-2e9f4c62	172.31.0.0/20

Cancel

Previous

Next: Review

2.23 Go to Target group and create a target group.

Create target group Actions

Filter by tags and attributes or search by keyword

<input type="checkbox"/>	Name	Port	Protocol	Target type	Load Balanc	VPC ID	Monitoring
<input type="checkbox"/>	serveroneatg	80	HTTP	Instance	netbankinglb	vpc-fa99e992	<input type="checkbox"/>
<input checked="" type="checkbox"/>	serveronebtg	80	HTTP	instance		vpc-fa99e992	<input type="checkbox"/>

2.24 Go to load balancer listener and select “view edit rules” and add rules.

RULE ID	IF (all match)	THEN
1 A rule ID (ARN) is generated when you save your rule.	<div>✎ Path is /serveroneb/*</div> <div>+ Add condition</div>	1. Forward to... serveroneb1g

2.25 Save the rules and copy the DNS and check.

2.26 If you stopped one new instance should start spinning

	serveronea	i-0e18a71a4b91e9569	t2.small	ap-south-1b	stopping	Load Balancing
	serveroneb	i-0e18a77828a3d879c	t2.small	ap-south-1b	running	2/2 checks... Load Balancing

	new instance	i-04b2a499348213e00	t2.small	ap-south-1a	pending	Initializing
	serveroneb	i-0e18a77828a3d879c	t2.small	ap-south-1b	running	2/2 checks... Load Balancing