

Given the above reference architecture, please explain the areas of configuration, monitoring, and troubleshooting that would have to take place in the following scenarios, across both configurations:

1. How the networking and load balancing should be configured
2. What monitoring and feedback systems you would put in place to aid in troubleshooting potential issues
3. Specific troubleshooting steps you would take to arrive at a solution for each error symptom (perhaps in deicision tree format)

Please articulate the troubleshooting steps (in detail) and tools you would use to resolve.

###### How should networking and Load Balancing be configured?

**Application LB:**

* Configure at least one public subnet for each of the availability zones
* Create and configure an application LB
  + Give the LB a name
  + For Availability Zones (az) choose the VPC for the EC2 instances hosting the application, and the public IP for each az
  + Configure HTTPS
* Create and Configure Security Groups (sg) for the LB
  + Create a new sg
  + Configure sg routing
* Create and Configure Target Group (tg)
  + Give the TG a name, protocol HTTPS, Target Type = instance, accept the default health check
* Register targets with the TG
  + Assuming we have compute instances, select each instance that will be part of the TG
* Once the Application LB is created, test it

**Classic ELB:**

* Define the VPC and Subnets, ensure you have a CIDR block of at least a /27 (e.g. 192.0.0.0/27) to scale your application
* Create a private subnet in each AZ for your compute instances, and public subnets for the LB
* Create Security groups for the instances that allow traffic in both directions and ports for each subnet
* Create and configure Network access control lists (ACL) – the default ACL may or may not be sufficient
* Launch instances that will register with the ELB, make sure instance sg allows HTTP traffic on port 443
* Create and define the ELB
* Name the ELB, and configure any advanced VPC configuration(s)
* Either leave the default listener config or modify it as needed (e.g. HTTPS/PORT 443)
* For the EC2 VPC available subnets configure at least one of your available public subnets, adding one per AZ as you need
* Assign the default SG for the Classic EC2 LB, as needed assign custom SG (new or existing), configure the listener to be secure if required
* Setup and configure health checks
* Register instances with the LB

###### What monitoring and feedback systems you would put in place to aid in troubleshooting potential issues

**Application LB:**

In specific terms for the ALB itself I’d start with AWS’s own tooling such as:

* CloudWatch metrics
* Request Tracing
* Access Logs
* CloudTrail Logs

If these aren’t sufficient, it may be prudent to look into other management, monitoring, and alerting solutions; e.g. New Relic Data Dog, Digg

**Classic ELB:**

* CloudWatch metrics
* ELB Access Logs
* CloudTrail logs

###### Specific troubleshooting steps you would take to arrive at a solution for each error symptom

In the included diagram I don’t see any error conditions, I’ll walk through a few common error situations

Example Error Symptoms:

* HTTP error 503: Service Unavailable:
* Clients unable to connect to your public facing LB
* Health Checks failing

Please articulate the troubleshooting steps (in detail) and tools you would use to resolve.

**Application LB:**

For the *Error 503* the problem is almost certainly there are no targets registered with the configured LB target group(s). Simply check your target groups and verify there are compute instances configured for the TG and go back to the Application LB config and select (or create if that was missed somehow) the compute instances / targets for the LB target group. Create targets if this was missed.

**Classic ELB:**

For the Classic ELB there may be one of several reasons for this:

* No healthy instances
* No registered instances
* Insufficient Capacity in the Classic ELB

To troubleshoot:

1. **No healthy instances**: In the CloudWatch metrics check the ‘HealthyHostCount’ metric. You’ll need at least one healthy instance per AZ. Can you curl the endpoint If needed, start new compute instances?
2. **No registered instances**: Ensure you’ve registered at least one instance per AZ the LB is configured to respond. Check the CloudWatch ‘HealthyHostCount’ metric. Check to see if you should enable and configure cross-zone load balancing if you’re unable to ensure a registered instance in each AZ
3. **Insufficient Capacity**: This is most often a transient issue that resolves itself. Anything >5 minutes you should open a support request with Amazon.

*Clients unable to connect to your public facing LB*

**Application LB:**

If the App LB is not responding check:

1. Is the LB only attached to a private subnet? If so you have to specify public subnets for the LB, and this public subnet has to rout to your internet gateway for the configured VPC. If this is configured properly check:
2. Security Groups and Network ACLs: The SG or possibly the Network ACLs for the LB subnet isn’t configured to allow inbound traffic from clients and outbound traffic to clients over the listener port(s). If either or both are the case, configure the SG and any associated network ACL to allow the inbound and outbound traffic

**Classic ELB: T**he Classic ELB and the App LB have the same troubleshooting steps for this error condition

1. Is the ELB attached to a private subnet? If so you have to specify public subnets for the LB, and this public subnet has to rout to your internet gateway for the configured VPC
2. Just like the App LB Security Groups and Network ACLs could be the culprit if the subnet is configured properly: The SG or possibly the Network ACLs for the LB subnet isn’t configured to allow inbound traffic from clients and outbound traffic to clients over the listener port(s). If either or both are the case, configure the SG and any associated network ACL to allow the inbound and outbound traffic

*Health Checks failing*

**Application LB:**

1. Is the webserver running in the instance(s) behind the LB running?
   1. Yes: move to step 2
   2. No: check the configuration of your webserver
2. Can you curl the healthcheck.html endpoint, and what is the response? (200 OK is what we need for the health check to pass)?
   1. yes: move to step 3
   2. No: Is the instance running or move to step 3, perhaps a blocked port or ports
3. Does the instance have a SG which permits access over port 443 (or 80)?
   1. yes: All the Health checks should now be passing
   2. no: Configure the instance SG to allow access over port 443 and test.
      1. Sometimes you may want to open all ports and once you’ve got it working, then configure specific ports and retest

**Classic ELB:**

Failing Health Checks in and troubleshooting for the classic ELB have some of the same basic troubleshooting and root-causes. Tools and utils you’ll need to use as part of the troubleshooting process include curl, ssh, tcpdump, and netstat.