AWS Solution Architect
Service unavailable report

Niall Creech

October 3, 2019

1 Solution analysis and fixes

The current implementation cannot be accessed from external networks. The service consists of,

- Networking is a VPC with CIDR 10.0.0.0/16 with 2 public subnets 10.0.0.0/24 and 10.0.1.0/24 in eu-west-1b and eu-west-1b respectively, as well as 2 private subnets 10.0.2.0/24, 10.0.3.0/24 in eu-west-1b and eu-west-1b.
- VPC has an internet gateway and a classic load balancer (ELB)
- Routing is setup up to allow all traffic between subnets and out to external addresses.
- There are 2 security groups, one for the single EC2 application instance, and one for the ELB.
- DNS is enabled and the instane has a pubic IP address

1.1 ELB healthcheck on wrong instance port

Resolution: Change ELB port for healthcheck from 443 to 80
Detail: Httpd starts on-instance on port 80. The ELB correctly forwards ports from from 80 to instance port 80. However, the healthcheck looks to verify instance health on port 443 through a tcp connection. This change simply targets the healthcheck correctly at port 80 on the instance.

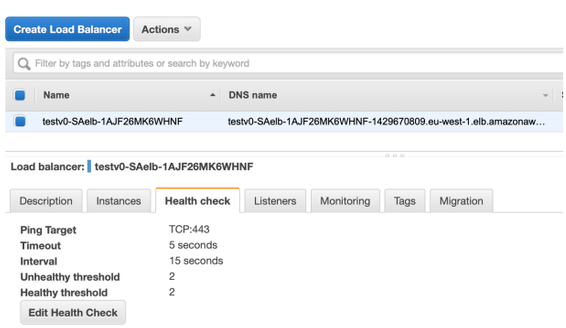

Figure 1: In the AWS console, navigate to Services → EC2 → Load Balancers

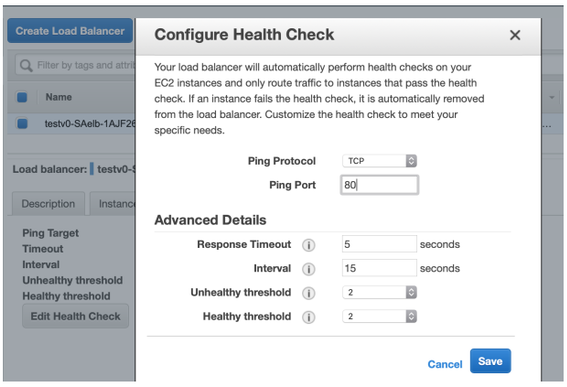

Figure 2: In the 'Healthchecks' tab, and change the port to 80

1.2 The site cannot be accessed from external addresses

Resolution: Add everywhere access to load balancer ingress
Detail: Allow all addresses to access the load balancer on port 80

1.3 Load balancer cannot route to subnet

Resolution: Add PublicSubnetA to load balancer
Detail: The ELB can only route to instances in PublicSubnetB. This change allows it to route to PublicSubnetA also, where the current instance is located.

1.4 Instances cannot be accessed from the load balancer

Resolution: Allow access to the instance from the ELB only
Detail: This change allows traffic from the ELB to the instances. This is set to all TCP traffic to port 80 to allow the ELBs TCP-based healthcheck to pass

2 Solution Enhancement

2.1 Operational

The ability to run and monitor systems to deliver business value and to continually improve supporting processes and procedures

2.1.1 Improve monitoring and logging

Monitoring and logging on the existing instance can be increased by enabling the cloudwatch agent on the application instances

2.2 Security

The ability to protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies

2.2.1 Harden EC2 instances

2.2.2 Enable SSM session manager

2.2.3 Strengthen NACL to prevent cross-subnet traffic

2.2.4 Only allow NACL to pass traffic to private subnets from public subnets

The private NACL allows all external traffic to pass. We can increase security by allowing only access from public subnets, or additionally from private endpoints, NAT gateways etc.

2.3 Reliability

The ability of a system to recover from infrastructure or service disruptions, dynamically acquire computing resources to meet demand, and mitigate disruptions such as misconfigurations or transient network issues

Currently there is a single instance with no auto-recovery

2.3.1 Replicate instances into multiple AZs

2.3.2 Create an auto-scaling group for application instances

2.4 Performance

The ability to use computing resources efficiently to meet system requirements, and to maintain that efficiency as demand changes and technologies evolve

2.4.1 Change instance type

The current solution uses bursting t2.micro instances with CPU credit limits. This is very cost efficient, but performance is not suitable for reasonable traffic levels and exhausting CPU credits will cause degradation of availability.

2.5 Cost

The ability to run systems to deliver business value at the lowest price point