CSAA Bootcamp – June 2019

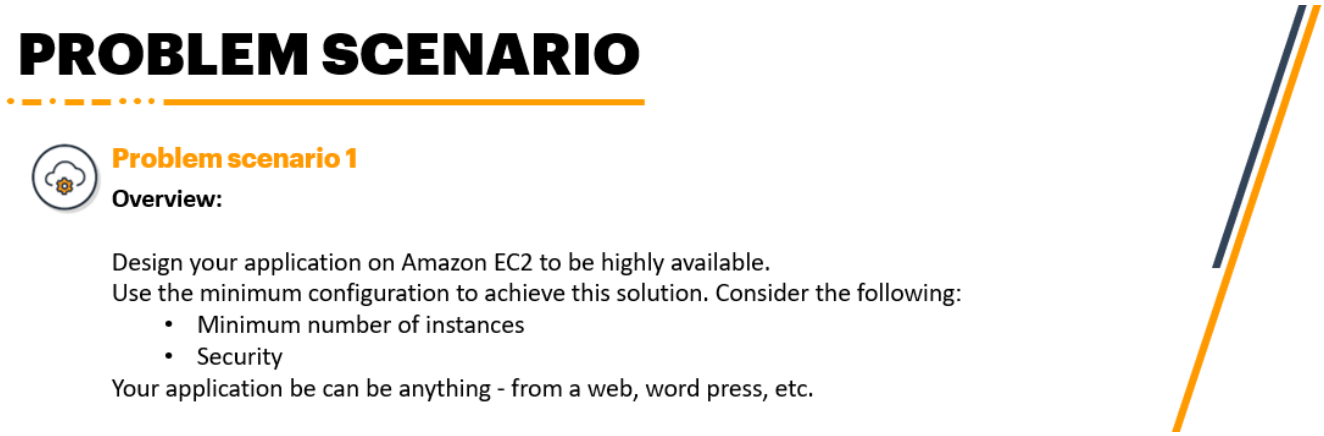
Problem Scenario #1

Table of Contents

1 DOCUMENT DEFINITION

The purpose of this document is to present a solution to the Problem Scenario #1.

2 PROBLEM STATEMENT



3 SOLUTION

3.1 Key Considerations

The following are the key considerations for this solution:

1. Minimum configuration in AWS
2. Minimum number of instances
3. Applied security
4. Highly Available

3.2 Technology Assumptions

The following are the technology assumptions for this solution:

1. Data persistence is not required.
2. Security only covers network security.
3. Failover for the EC2 instance is not required.
4. User-friendly FQDN is not required.
5. Application code is stored in a git repository.
6. Application code is runnable in any terminal.
7. VPC and IGW are present and configured with bidirectional traffic.
8. All incoming traffic are limited to HTTPS.
9. An entire AWS region will never fail. The application will be hosted in US East 1 (North Virginia).

10. At most one availability zone is expected.
11. Downtime is not acceptable. But a decrease in performance is acceptable.
12. Advanced monitoring is not required.
13. Cross-origin is not required.

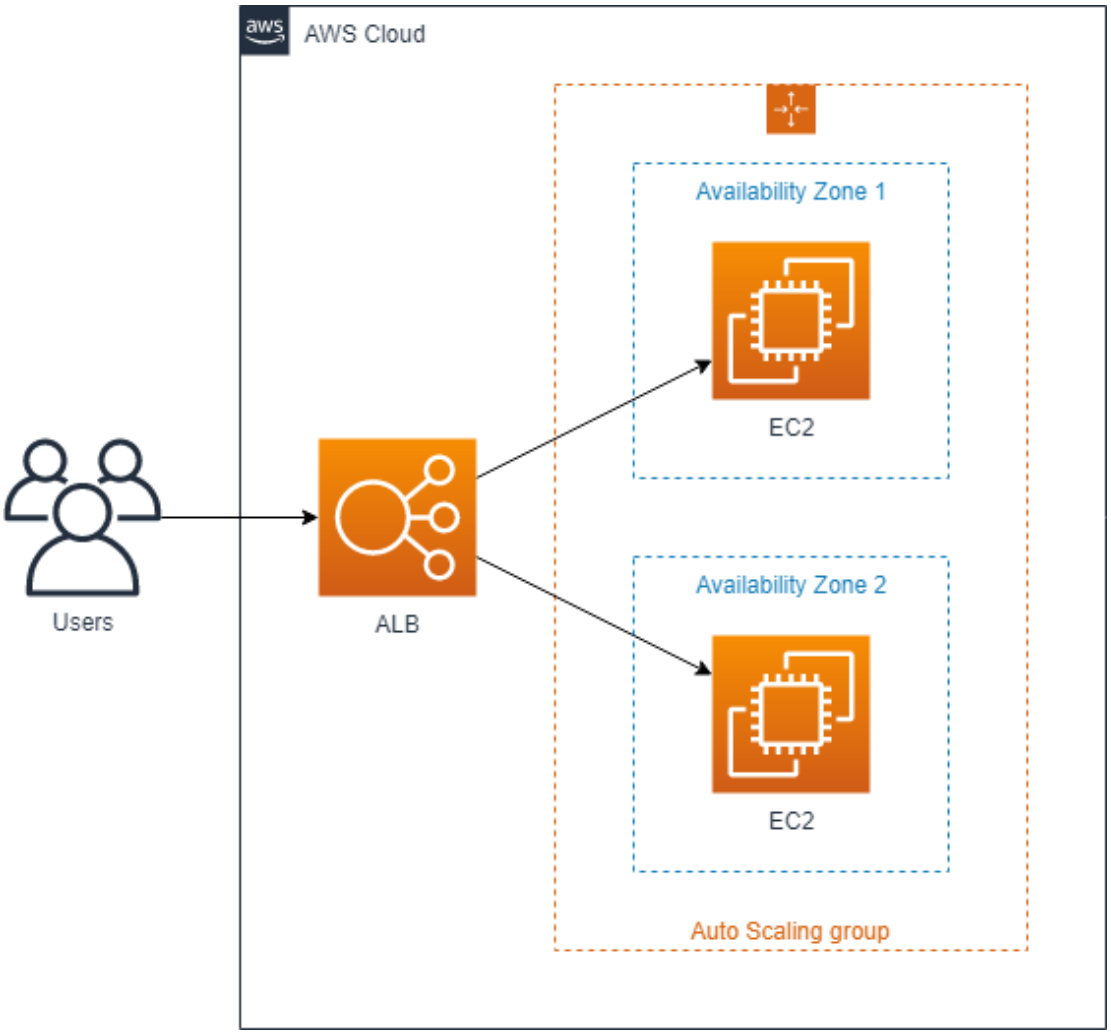
3.3 AWS Services

To achieve the considerations defined, the following AWS Services will be required:

1. Application Load Balancer (ALB) – distributes traffic across the target group that is application-aware
2. Elastic Compute Cloud (EC2) – IaaS-based compute service
    a. Security Group – configures ingress traffic coming to the EC2 instances
    b. Auto-scaling Group – enables automated scaling of EC2 instances

3.4 Architecture Design

Below is the architecture design to achieve the key considerations:



ALB is used to distribute traffic across the EC2 instances inside auto-scaling groups. It can also automatically target newly created instances inside an auto-scaling group (ref. Auto Scaling Benefits). ALB will also be configured with Cross Zone to support cases where an entire AZ becomes unavailable.

An auto-scaling group will be created with a minimum of one EC2 instance and at most three. A boot strap script will be ran in each newly created EC2 instance which will clone a private git repository containing the application code using personal token, build the application code and run the built application package in terminal.

The EC2 instances will be mapped to a single security group with only port 443 allowed for ingress traffic.