Problem Statement

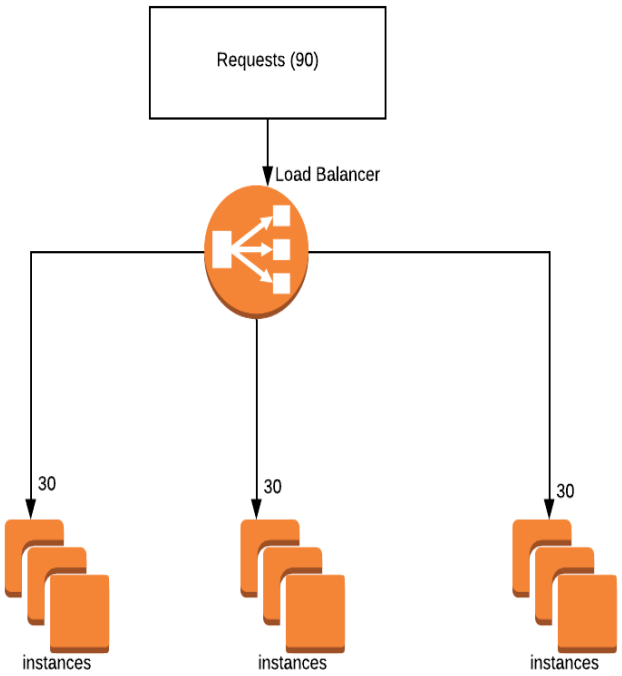
Assume a web service running out of AWS In order to configure HA with region failover for this service, It would need to be running from a minimum of two different regions. AWS LB does not support Load Balancing across regions End Points.

Load Balancer

A load balancer distributes incoming application traffic across multiple EC2 instances in multiple Availability Zones. This increases the fault tolerance of your applications. Elastic Load Balancing detects unhealthy instances and routes traffic only to healthy instances

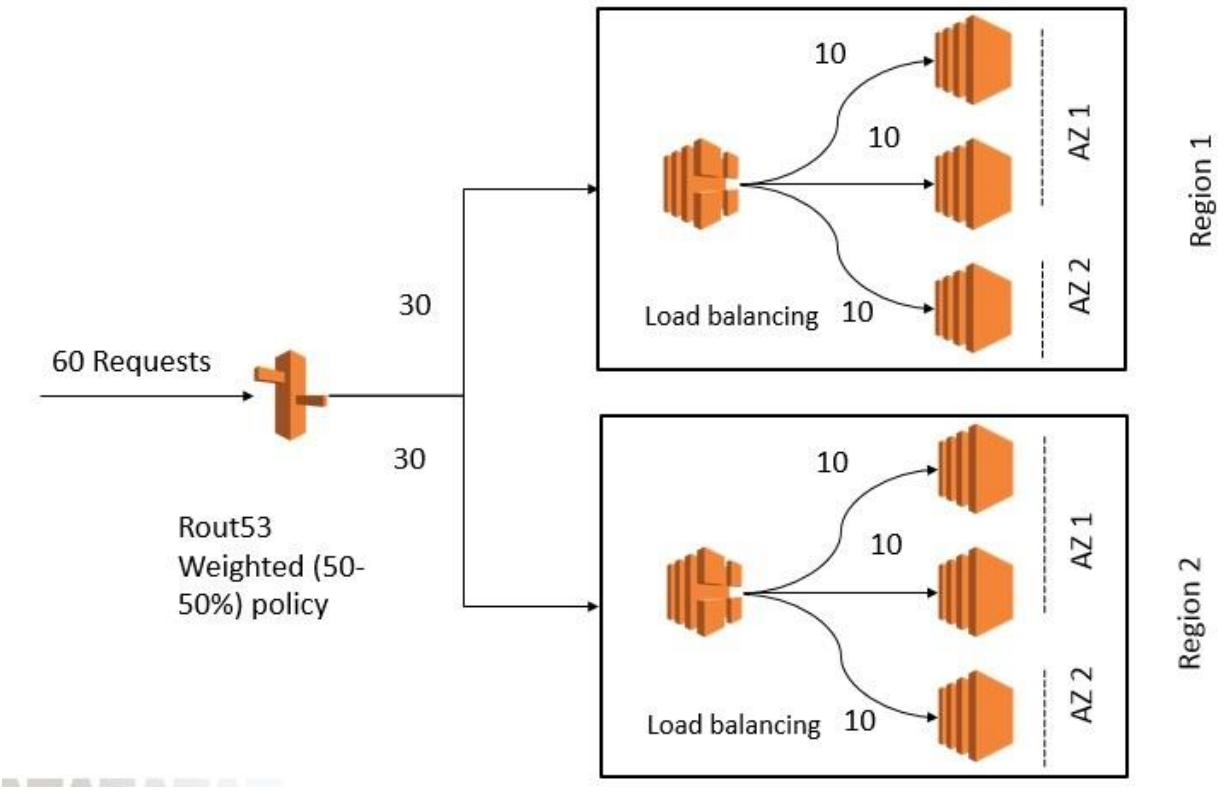
Load balancer serves as a single point of contact for clients. This increases the availability of your application. You can add and remove instances from your load balancer as your needs change, without disrupting the overall flow of requests to your application. Elastic Load Balancing scales your load balancer as traffic to your application changes over time. Elastic Load Balancing can scale to the vast majority of workloads automatically.

Load balancers are regional service i.e. they operate within an AWS region. Classic Load Balancers support for EC2 classsics.

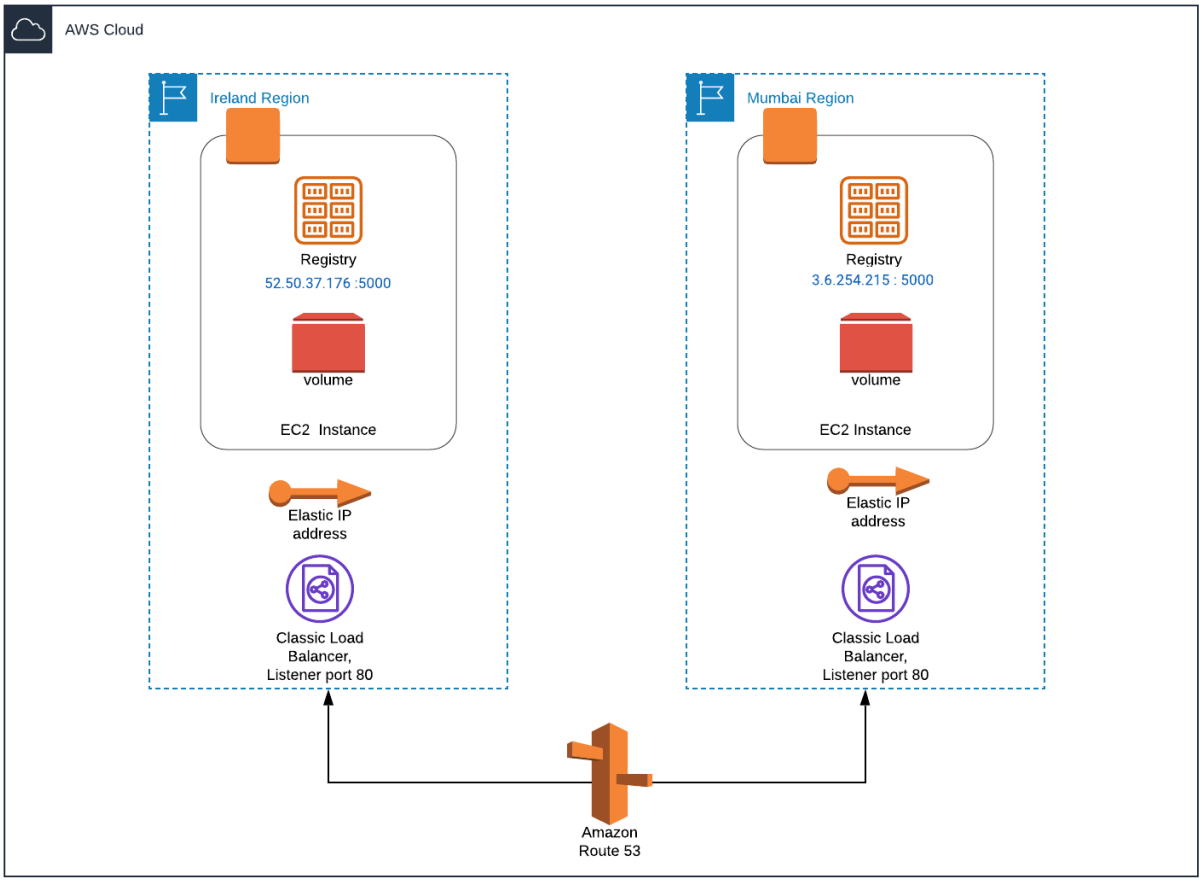


A *listener* checks for connection requests from clients, using the protocol and port that you configure, and forwards requests to one or more registered instances using the protocol and port number that you configure. You add one or more listeners to your load balancer.

Load Balancing between two services running on to Amazon Instances across two regions



Load balancing between two regions in an amazon Instance is effectively done using Route 53 routing policy. Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. It is designed to give developers and businesses an extremely reliable and cost effective way to route end users to Internet applications by translating names like www.example.com into the numeric IP addresses like 192.0.2.1 that computers use to connect to each other. Amazon Route 53 is fully compliant with IPv6 as well.



To do this, You would need to setup and Instance in 2 regions as shown in the above diagram Ireland is used as the first region and Mumbai is used as the second region. And then open port 5000 in both the regions. Install registry in both the instances, Check if the registry is running properly in port 5000, Setup Elastic IP's which are associated with both the Instances. Elastic IP is used so that the IP address of the instance remains constant. Create Load balancers associated with port 80. These load balancers must be added in Route 53 entry.

Route 53 effectively connects user requests to infrastructure running in AWS – such as Amazon EC2 instances, Elastic Load Balancing load balancers, or Amazon S3 buckets – and can also be used to route users to infrastructure outside of AWS. You can use Amazon Route 53 to configure DNS health checks to route traffic to healthy endpoints or to independently monitor the health of your application and its endpoint.