Case Study : Ecommerce

Background: A leading smart commerce platform in Asia, with offices in different locations that have help more than a quarter million merchants open their online stores. The business supports brands of all sizes to achieve local and international growth with omnichannel presence.

High Availability functions as a failure response mechanism for infrastructure and makes system recover from component failure automatically. The goal is to eliminate single points of failure in the infrastructure.

We adopt a top-to-bottom approach to implement failure detection and recovery for redundant systems (i.e. the layer on top becomes responsible for monitoring the layer immediately below it for failures). In our scenario (as shown in the diagram), the load balance is the top layer and the database master is the bottom layer.

1) Replication of database: if one of the database becomes corrupted for some reasons, say, due to AZ failure, the other database can be kicked in as a back up for ensuring data integrity.
2) Servers redundancy: We have an infrastructure consisting of two identical, redundant servers (EC2 auto scaling in our case), if one of servers goes down, the load balancer will redirect all traffic to the remaining online server.
3) Clustering of load balancers and failover: Active and passive load balancer are connected as a cluster where each node is capable of failure detection and recovery. Here recovering from a load balancer failure involves a failover to a redundant load balancer. With the aid of Reserved IP, we can keep the domain name associated with the same IP address, and only the IP address itself is moved between the nodes.
4) Geographic redundancy: here two servers are installed in two different AZs to achieve high availability so that if one location fails, the other can still function.
5) Software for configuration: HAProxy (High Availability Proxy) is a common choice for load balancing as it can handle load balancing and different kinds of servers.

As a conclusion, moving up in system stack, it is important to implement a reliable redundant solution for our application entry point, normally the load balancer. To remove this single point of failure, we need to implement a cluster of load balancers behind a Reserved IP. Lastly, HA firewall, typically multiple

WAF are placed throughout the network to help eliminate any single point of failure and enable ongoing failover processing.

EC2 instance high availability using auto scaling , replication of DB , geo-redundancy and configure additional load balancer to achieve redundancy using a floating IP address