AWS Route 53 Routing Policy

AWS Route 53 routing policy determines how AWS would respond to the DNS queries and provides multiple Routing policy options

Simple Routing Policy

* Simple routing policy is a simple round robin policy and can be applied when there is a single resource doing the function for the domain *for e.g. web server that serves content for the website*
* AWS Route 53 responds to the DNS queries based on the values in the resource record set *for e.g. ip address in an A record*

Weighted Routing Policy

* Weighted routing policy enables Route 53 to route traffic to different resources in specified proportions (weights) *for e.g., 75% one server and 25% to the* *other during a pilot release*
* Weights can be assigned any number from 0 to 255
* Weighted routing policy can be applied when there are multiple resources that perform the same function *for e.g., webservers serving the same site*
* Weighted resource record sets let you associate multiple resources with a single DNS name
* Common use cases include
  + load balancing
  + A/B testing and piloting new versions of software
* To create a group of weighted resource record sets, two or more resource record sets can be created that have the same combination of DNS name and type, and each resource record set is assigned a unique identifier and a relative weight.
* When processing a DNS query, Route 53 searches for a resource record set or a group of resource record sets that have the specified name and type.
* Route 53 selects one from the group. Probability of any one resource record set being selected depends on its weight as a proportion of the total weight for all resource record sets in the group *for e.g., suppose for www.example.com has three resource record sets with weights of 1 (20%), 1 (20%), and 3 (60%)(sum = 5). On average, Route 53 selects each of the first two resource record sets one-fifth of the time, and returns the third resource record set three-fifths of the time.*

Latency-based Routing (LBR) Policy

* Latency-based Routing Policy enables Route 53 to respond to the DNS query based on which data center gives the user the lowest network latency
* Latency-based routing policy can be used when there are multiple resources performing the same function and Route 53 needs to be configured to respond to the DNS queries with the resources that provide the fastest response with lowest latency
* Latency resource record set can be created for the EC2 resource in each region that hosts the application. When Route 53 receives a query for the corresponding domain, it selects the latency resource record set for the EC2 region that gives the user the lowest latency. Route 53 then responds with the value associated with that resource record set *for e.g., you might have web servers for example.com in the EC2 data centers in Ireland and in Tokyo. When a user browses to example.com from Singapore, Route 53 will pick up the data center (Tokyo) which has the lowest latency from the users location*
* **Latency between hosts on the Internet can change over time as a result of changes in network connectivity and routing. Latency-based routing is based on latency measurements performed over a period of time, and the measurements reflect these changes***for e.g. if the latency from the user in Singapore to Ireland improves, the user can be routed to Ireland*
* Latency based routing cannot guarantee users from the same geographic will be served from the same location for any compliance reason
* Latency resource record sets can be created using any record type that Route 53 supports except NS or SOA

Failover Routing Policy

* Failover routing policy allows active-passive failover configuration, in which one resource takes all traffic when it’s available and the other resource takes all traffic when the first resource isn’t available.
* Route 53 health checking agents will monitor each location/endpoint of the application to determine the availability.
* Failover routing policy is applicable for Public hosted zones only

Geolocation Routing Policy

* Geolocation routing policy enables Route 53 to respond to DNS queries based on the geographic location of the users i.e. location from which the DNS queries originate
* Common use cases include
  + localization of content and presenting some or all of the website in the users language
  + restrict distribution of content to only the locations in which you have distribution rights.
  + balancing load across endpoints in a predictable, easy-to-manage way, so that each user location is consistently routed to the same endpoint.
* Geolocation routing policy allows geographic locations to be specified by continent, country, or by state (only in US)
* Geolocation record sets, if created, for overlapping geographic regions *for e.g. continent and then for the country within the same continent*, priority goes to the smallest geographic region, which allows some queries for a continent to be routed to one resource and queries for selected countries on that continent to a different resource
* Geolocation works by mapping IP addresses to locations, which might not mapped to a exact geographic location
* A default resource record set can be created to handle these queries and also the ones which do not have an explicit record set created
* Route 53 returns a “no answer”response for queries from those locations, if a default resource record set if not created
* Two geolocation resource record sets that specify the same geographic location **cannot** be created
* Route 53 supports the edns-client-subnet extension of EDNS0 (EDNS0 adds several optional extensions to the DNS protocol.) to improve the accuracy of geolocation routing