Route53

Uday Manchanda

August 8, 2021

1 Route53

- Managed DNS
- the most common records are:
  - A: hostname to IPv4
  - AAAA: hostname to IPv6
  - CNAME: hostname to hostname
  - Alias: hostname to AWS resource
- Diagram for A record
  - Browser sends DNS request to Route53 with hostname
  - Route53 responds with IP
  - Browser makes HTTP request to the IP address, application server responds with HTTP response
- Can use public (you own or buy) or private domains (resolved by your instances in your VPCs)
- Load balancing
- Health checks
- Routing policy

1.1 TTL

- Time to live, its a DNS record
- cache the response of a DNS query, don't want to overload DNS
- As soon as the browser sends a DNS request, it will cache it, along with the response
- Higher TTL means less traffic on DNS but possibility of outdated records

1.2 CNAME vs Alias

- AWS resources expose a hostname, but you want to expose a url
- CNAME: points a hostname to any other hostname, only for non-root domain (ex: something.domain.com)
- Alias: points a hostname to an aws resource (app.mydomain.com → blabla.domain.com). Free of charge, native health check
- If you have a root domain use alias, if it's not a root domain you can use either (alias preferred)

1.3 Health Checks

- if an instance is unhealthy, R53 will not send traffic to that instance
- instance is unhealthy if fails three HCs in a row, opposite for healthy
- default HC interval is 30 seconds
- HTTP/TCP/HTTPS health checks
- can integrate with CW and linked to R53 DNS queries

2 Routing Policy

- Simple
  - use to redirect to a single resource
  - no health checks
  - if multiple values are returned, a random record is returned
- Weighted
  - controls the percentage of the requests that go to a specific endpoint
  - EX: three EC2 instances. R53 sends x percent to EC2A, y percent to EC2B, and z percent to EC2C
  - helpful to test 1 percent of traffic on new app version for exmaple
  - Helpful to split traffic between regions
  - Will create a number of A records depending on how many IP addresses are listed
- Latency
  - Redirect to the server that has the least latency close to us
  - Latency is evaluated in terms of user to designated AWS region

- Failover
  - Two EC2 instances, primary and secondary
  - HC on the primary instance, if it fails, it will failover to the secondary
- Geolocation
  - Different from latency, routing based on user location
  - here we specify: traffic from the UK should go to this specific IP
  - Should create a default policy in case there's no match on the location
- Geoproximity
  - Route traffic to your resources based on geographic location of users and resources
  - Ability to shift more traffic to resources based on the defined bias
  - Reosurces can be AWS (like AWS region) or not (like Lat/Long)
  - Specify bias values to change the size of the geographic location
  - Useful if you need to shift resources to a certain region or vice versa
- Multivalue
  - Route traffic to multiple resources
  - Want to associate a R53 HC with records
  - Up to 8 healthy records are returned for each multivalue query
  - Not a subsitute for ELB