

Welcome back and in this lesson, I want to talk about latency based routing within Route 53. So this is going to be a relatively brief lesson. I just want you to be aware of the architecture of how latency based routing works because you'll need it for the exam. Now, the premise of latency based routing is pretty simple to understand. Essentially, for every record that you create that uses latency based routing you specify a region. Now, let me show you how that works. If I go to hosted zones and then go into associatecats.com, which is the hosted zone that I've been using throughout this topic, I've already got these three individual records. So I'm going to go ahead and delete these. Once I've done that, I'm going to create a record set I'm going to call it www I'll need the IP address of the web1 instance. So I'm going to move to the EC2 console, copy that into my clipboard, and then paste that in as the value and set the one minute TTL and then once I've done that, I'm going to change the routing policy from simple to latency and by default, it set AZ U.S. East One as the region. Every single time you create a latency based routing policy record, you need to specify the region. Now, all of these EC2 instances that I've got created are in the U.S. East 1 or North Virginia region but for the sake of this demonstration, I want you to pretend that they're not. I'm going to create this first record, and I'm going to set this to U.S. East 1 I'll need to configure a set ID that's unique amongst all of these records with the same name. So I'm going to use the same thing. I'm going to say U.S.-East-1 and they should just remind me that this is based in the U.S. East 1 region. So I've done that. I'll hit create. Now, once I've configured this record, I'm going to create a second one. So I'm going to create record set www again, one minute TTL, move back across to the EC2 to console. I'm going to select web2 copy that IP address into my clipboard, paste that into the value, change the routing policy down to latency. Now I'm going to change the region from U.S. East 1 to E.U. Central 1 to match the diagram on the right of my screen. So I'm going to scroll down to E.U. Central 1 and select that I will get a warning because, of course, the EC2 instance that I'm pointing out is in the U.S. East 1 and so where Route 53 is smart enough to warn us of that fact. What we would do if this was production, will be to create these EC2 instances in the same region that we're setting in this dropdown but it's fine for this demonstration. I'll just go ahead and ignore that and leave it set to E.U. Central 1. Now go to configure the set ID to be the same. So E.U. Central-1 and I click on create. Now, in addition to the ones that are on my diagram here, I'm going to go ahead and create a final record set and we call this one www, I'm going to use a one minute TTL. I'm going to select the web three instance and copy that into my clipboard, paste this in, change it from simple to latency this time, however, I'm going to select a A.P. Southeast 2 which is my home region. So I select that I'll get the same warning. We don't need to worry about it because it's just warning us that the resource is not in this region and I'll call the set ID the same thing. So A.P.-Southeast-2 and that looks good, so I'll create that record. So now we've got three records. We've got one record. So web one is in U.S. East 1, web2 is in E.U. Central 1 and then, in addition to this diagram, I've got web3 which is an AP Southeast 2. Now the way that Route 53 works is that it maintains a database of latencies between internet based endpoints. So Route 53 knows if I try to resolve from London, it will know what the latency is between London and each of the different AWS regions and if you've got a correctly configured Route 53 latency based set of records, it will return the record with the lowest latency. So in this example, if I use a client in London and attempt to look up www.associatecats.com it will, in all likelihood return the record that is set to the E.U. Central 1 region because that offers the lowest latency. If the client were to be in America, it would probably return that U.S. East 1 record because that would offer the lower latency. Now, because I'm based in Australia. If I attempt perform a look up against this record set so www.associatecats.com it should, in theory, return the one that's closest to me, which is A.P. Southeast 2. So let's test this now if I perform a look up against this set of records because I'm based in A.P. Southeast 2 which is Australia, the value that I'm expecting to be returned is the value for this record set. So the one that's logged against the A.P. Southeast 2 region so that's the IP address that ends in 122.193. So I'm going to move across to my terminal and test that out. So I'll clear the screen to make it easier to see. I'll do a NS look up and then www.associatecats.com and then we got to get the 122.193 and that represents the lowest latency. It's the resource that's logged to the region. That's closest to me. So Route 53 knows that because I'm resolving these from Australia the lowest latency resource that matches the name that I'm trying to resolve is the one that's based in A.P. Southeast 2. So what I'll do now is I'm going to go ahead and delete this record set the one that's in A.P. Southeast 2 and just see which one is returned by Route 53 then so the only options that I have at this point are U.S. Central 1 and U.S. East 1. Now, these are regions of both based a fair distance from my current location. Now, because the U.S. East 1 is on the east coast of the U.S. and E.U. Central 1 is based in Europe we should find that both these are for a fairly similar latency for my current location, but I'd be interested to see which of them is returned. So If I go back to my terminal, I clear the screen, and I attempt to perform another look up on www.associatecats.com. This time, 55.166 is returned, and that's the one that's located in U.S. East 1. So from a latency perspective, Route 53 has determined that from my location to U.S. East one offers a lower latency than my location to E.U. Central 1 and that's the benefit that latency based routing provides. It allows you to direct DNS requests at the closest geographical region based on latency. Now from an exam perspective, **the critical thing that you need to understand is it is not only based on geography, it makes sense that the lowest latency from my location to any AWS region will be the region that is in A.P. Southeast 2. So if I'm doing a ping from my location, to any AWS regional endpoint the lowest latency will be the Sydney region but that's not always the case. If you're based in the U.S. it might be that your ISP has better latency to a region that potentially is a further distance away from a geographic perspective so latency based routing is not based on geography or distance. It's based on latency and whilst in most cases they are the same thing that isn't always the case. So with latency based routing, it's based on the conditions between you and the different AWS regions. It's not a distance thing. It's latency based so you can create individual records** and you can set them to be a specific region and when you do that, this database is consulted Route 53 selects from each of the individual regions that you've got tagged for this particular record set the one with the lowest latency and that's the one that's returned. Okay, so it's everything I wanted to cover in this lesson. I wanted to give you the basic architecture of latency based routing and explain that it's based on network conditions, not geography. In the next lesson, you're going to be looking at a different type of routing policy, which is based on geography but for the exam I want to make sure that you understand the difference, it's important because you might face exam questions where you need to provide a customer or customers with the best network performance and generally you do that using latency based routing not geolocation based routing they're different things and have different ideal use cases. So in the next lesson I will be talking about geolocation routing, which offers a way to direct traffic to the same geographic region that your customers are based in and the use cases for that very different. So go ahead, mark this lesson as complete, and when you're ready, you can join me in the next.