**EC2 Instance Types**

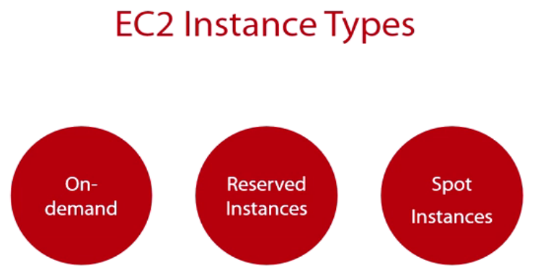
**Module Introduction**

**Well you can't build a building without knowing all the building blocks, now can you**? You can't build an infrastructure.

You can't build an environment on AWS without knowing all of the components, especially the compute component portion.

**Understanding EC2 Instance Types**

It's important to understand the different EC2 instance types before you can make a proper selection.



So you need to know that there are three.

**On-demand**: which is the traditional one that you're used to.

That's where you go in, you launch an instance, and you're basically paying by the hour for what you use.

But that's the default one, kind of the on-demand one.

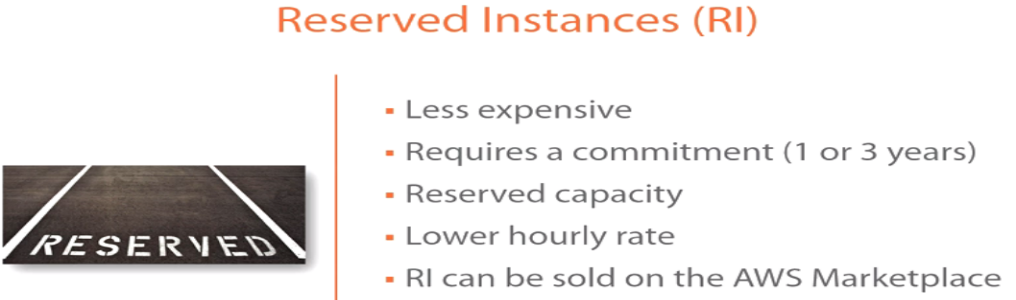
**Reserved instance:** A reserved instance is essentially the same as an on-demand instance except it has commit characteristics, which means you are committing to using this machine, this instance for an x amount of time.

In return, you're getting a discount for using it. You'll pay some upfront fee. Maybe you won't pay an upfront fee, but you're making a commitment as far as using this for a year or three years. And in exchange, you're getting a lower hourly rate. We're going to break that one down as well.

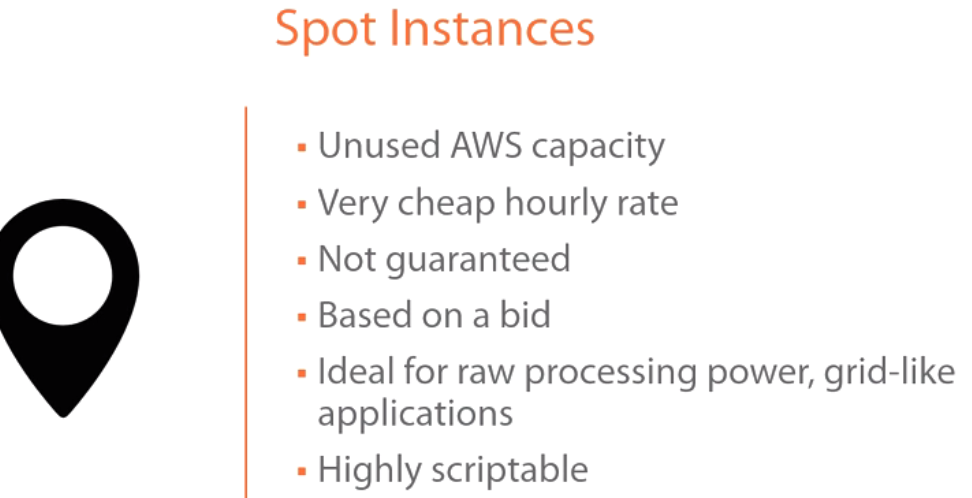
**Spot instances:** A spot instance is essentially a disposable machine. You bid on it, and you use it until somebody outbids you. Now it has its uses, and we're going to discuss those.

**On-demand instances:**

* Well, first, it's the default type. Whenever you go into the Management Console and you're provisioning an instance, it's the on-demand instance.
* It's the most expensive option. **And we're going to break it down.**
* There is no commitment. There's no upfront fee. You're just basically using it by the hour, and
* The prices are going to vary by AWS region, so North America might not be the same as your upper Asia, etc.
* It's billed on an hourly basis... 

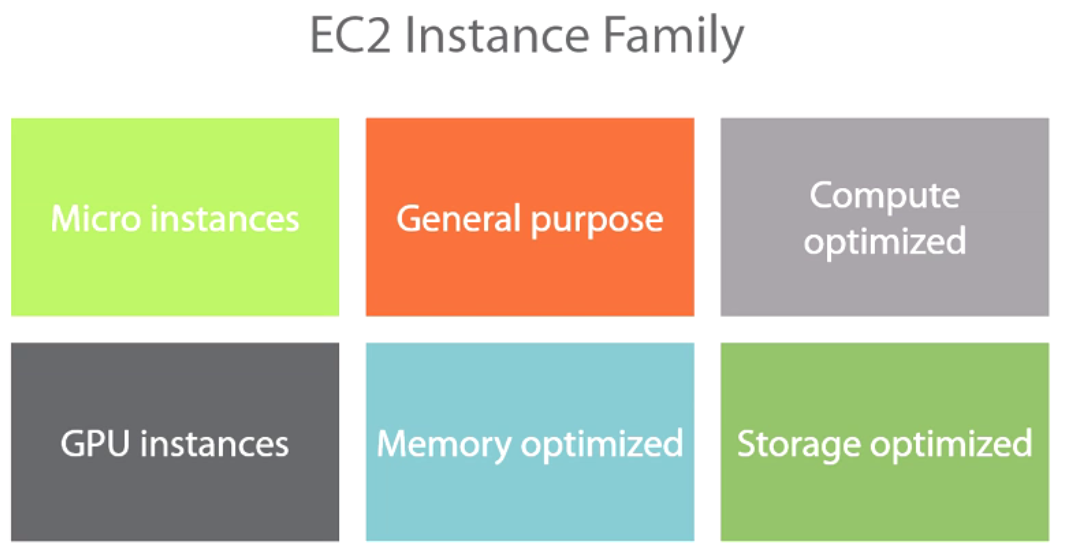
**A reserved instance**:

* Is less expensive.
* However, that less expensive requires a commitment. That commitment is either a 1- or 3-year commitment, so there is no 2-year. It's either a 1- or 3-year commitment. You're going to get a reserved capacity. So the advantage of a reserved instance not only is it going to give you a lower hourly rate because you're making this 1- or 3-year commitment.
* It’s also reserving the capacity, so it's ensuring if there is some run on an AWS availability zone, your capacity is reserved. AWS will never overprovision their availability zones and their data centers. As a result, when you're doing reserved instances, you are getting reserved capacity that machine will power on.
* You're getting a lower hourly rate.
* The reserved instances can also be sold on the AWS Marketplace.
* You're also committing to utilization with an RI.

**Now spot instances:** 

* These are available because of unused AWS capacity. So because AWS does not overcommit, they tend to end up with unused capacity, for example, at the end of the month or the end of the week, etc. This unused capacity, one of the ways that they put it to use, one of the ways that they monetize it is through the use of a spot instance
* . It's a very, very cheap hourly rate, and the reason it's cheap is, again, it's disposable. You can get outbid. If you get outbid, that means you're immediately going to lose access to this instance and somebody else is going to get access to it. So there's no notification, nothing.
* So it's not guaranteed. It's based on a bid, and
* It’s ideal for new processing power grid-like applications, maybe if you're mining for bitcoin, things that you can use on a mass scale. And if you lose nodes or instances, it doesn't really matter. That would be an ideal use case for a spot instance.
* It's highly scriptable, also, so you can spin up the spot instances based on a script, and you can get a lot of them, and you can do that in a very quick way.

**EC2 instance family:** Now when you go to provision an EC2 instance, there are different types of families that you're looking at.

An instance family is essentially a grouping of resources. In the general purpose, it's a little balance between CPU, memory, and disk. 

So for starters, you have **micro instances**. That's family---that's the lowest, and you're **not** going to get very good performance out of it. It's not very predictable, so that's the low-end. Typically, you're going to get this with the free tier.

**General purpose:** General purpose is a lot more stable. It gives you more memory, CPU configuration so you get a better configuration.

**Compute optimized**: it skews in the favor of compute, so it's a lot more compute heavy as opposed to being anything else, as opposed to being memory heavy.

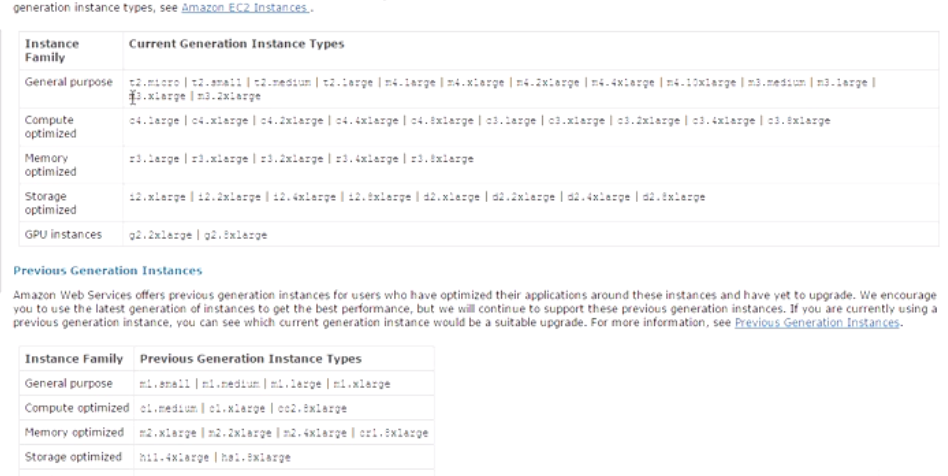
**GPU instances**. Those are going to have a graphical processing unit in it. And those can be used for specific workload, specific use cases.

**Memory optimized** machines that are heavy on memory, maybe less on CPU

**Storage optimized virtual instances** as well that are going to be heavy on the storage and less on other resources.

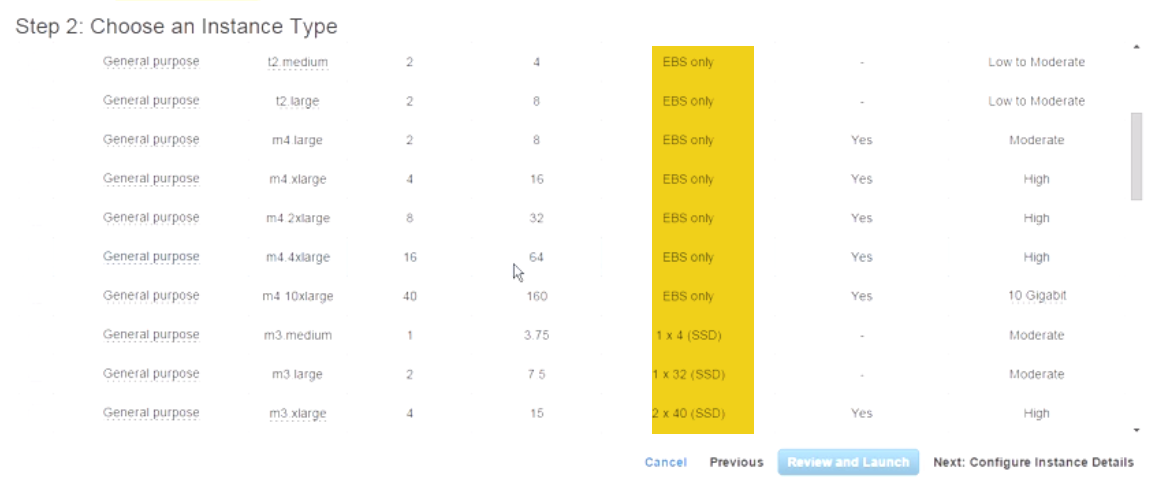
**Sizes and different configurations.**

So the general purpose---in general purpose, you're going to have two types of families. You're going to have the t, and you're going to have the m families.

This is the micro, again the cheapest level, typically the free tier. You're not going to get very good performance out of it. And then you have the m tier. And within the m family, the m3 family, you're going to have x1, x2, x3, so you're going to have multiple levels of configurations. This applies to the compute.

c3 means compute optimized, an r3 means memory optimized, an i2 means storage optimized, and so on and so forth, g2 means GPU optimized.

**EC2 Instance Types Walkthrough in the AWS Management Console**

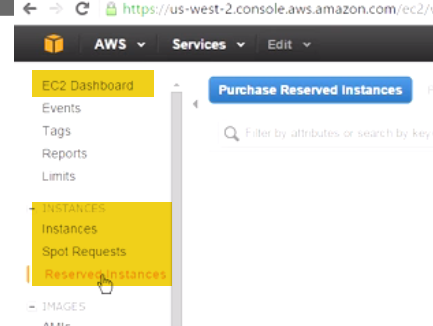
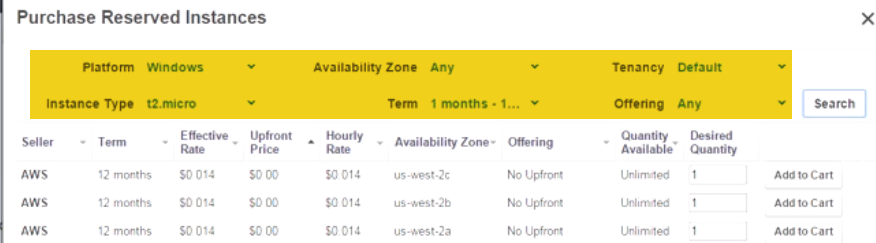
Management Console-> on-demand instance (Default)-> click on Launch Instance-> Windows Server 2012 R2->Down here you're going to get all of the instance families and their configurations, 

For example, the free tier, this one is the general purpose again. It's **a t2 micro. You're getting 1 vCPU. You're getting essentially 1 GB of memory** and so on and so forth. It's EBS, elastic block storage.

m4 here: two virtual CPUs, 8 GB of memory. It's EBS again and so on and so forth.

You can keep going down. This one is compute optimized, so you're getting a lot more compute here. You're still getting a ton of memory, but you see what I mean. And you can keep going down. I mean this one is giving you a ton of memory. You're getting 244 GB of memory here. So you can see how this works. The storage, again, they're giving you SSD here.

**Reserved instances**.

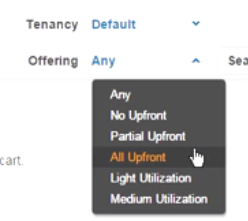
 

Purchasing a reserved instance is easy.

Click on Purchase Reserved Instance.

You're going to select the parameters to figure out what the pricing looks like.

* So, for example, if we wanted to select one that was Windows, and you see there're different options in here. You can **select Platform Windows.**
* **The instance type,** for example, we're buying this for a t2 micro. If you were buying it for a different instance that you have already provisioned, then you would select that from the list here, and you have all of the families that are listed,
* **Availability zone.**
* **From a term perspective:** recall I said you could only do either 12 months or 36 months, so one year or three years. Let's go ahead and select one year, for example.
* **From a tenancy perspective**, if you wanted to do hardware or dedicated, you could do that. I'm just going to keep it at default.
* **Offering perspective:**



I'm just going to keep it at any just to get some numbers here.

* And when you're ready with the criteria, you can click on **Search**, and it's going to give you a list of what you can purchase from a reserved instance for this and then apply it to an instance.

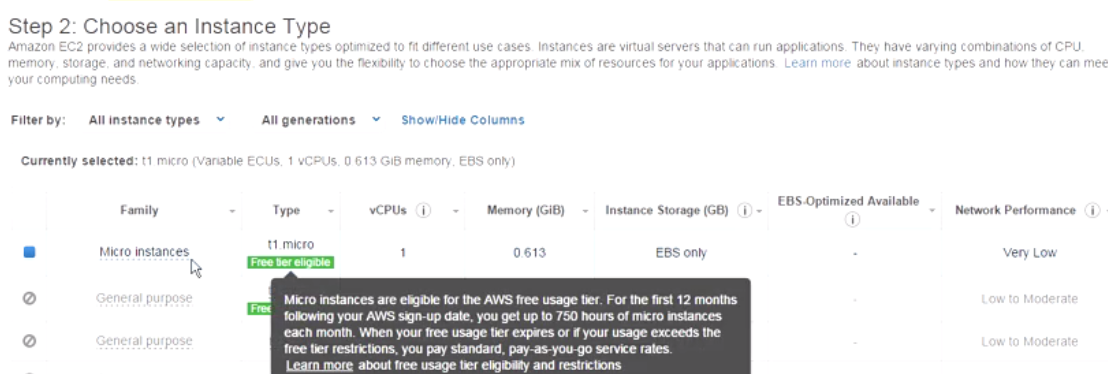
It's giving you the upfront price and how it varies if you put some money down. You can start to see how it varies from an effective rate perspective. You're also seeing the hourly rate, and you're seeing how that gets fluctuated based, again, on how much you put down. This is the availability zone.

**When do you use a reserved instance?**

You use a reserved instance when you know that you're already going to use this particular instance for a year. If you're going to use it for a year anyway, might as well **purchase a reserved instance for it because you're going to lower your hourly rate.** And then based on the utilization--light or medium or high--then if you're going to have it powered on all day, are you going to power down sometimes?, etc.

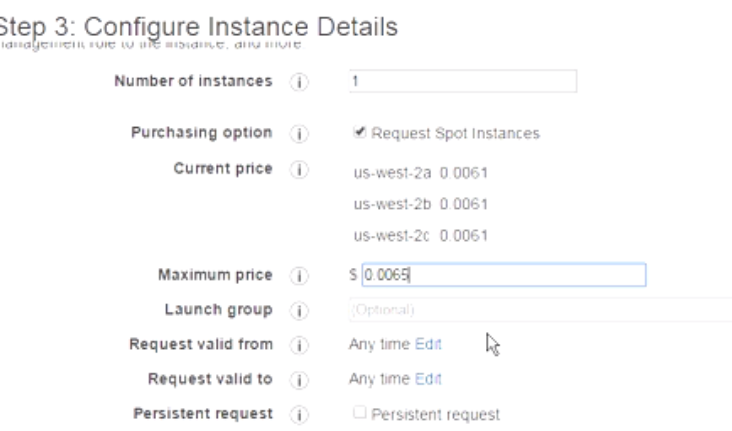
I don't recommend that you purchase the three-year term, although the three-year term will give you the highest level of discount possible. And the reason I don't recommend it is because if you look historically at how many types AWS has cut prices, if you're locked into a reserved instance and they happen to cut prices, then you don't take advantage of that.

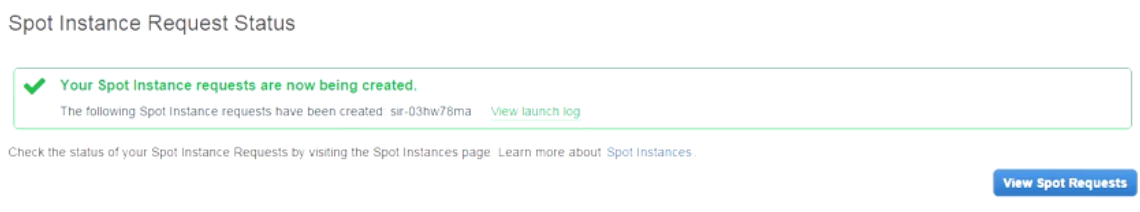
**Spot instances:** click on Spot Requests🡪click on Request Spot Instance, then you're going to select the (AMI.)Windows Server 2012, 🡪( instance family). What are you looking to figure---to bid on? I'm going to go with the free one just to keep things simple.

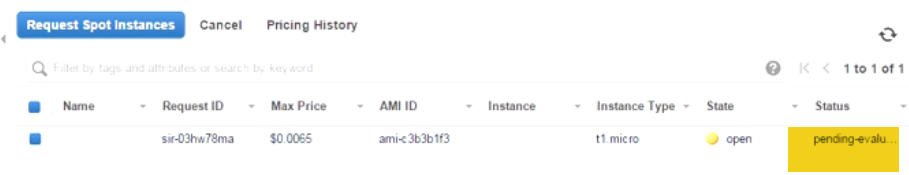
 Next to configure it. You go through the configuration, and here what you're going to see is:

How many do you want? How many spot instances to you want? : 1.

* You're requesting a spot instance.
* This is the current price.
* Now, what is the maximum price? What is the maximum bid that you're willing to pay for this instance? So maybe in my case I will do 0.0065 here just to kind of outbid the minimum bid.
* Again if you wanted to be part of a group, then you're saying that I want all of the instances that I'm provisioning to launch at the same time because I'm only provisioning one, this isn't going to apply to me at this point.
* The request is valid from, you can specify, and the request is valid to, you can also specify.
* And you can make this a persistent request, which means every time you fulfil this, every time the bid is--- the maximum bid is this, it's automatically going to submit a request for spot instances.
* So you can have this kind of in a repetitive format.
* The network that you want to put this in, again, which VPC? I'm just going to keep it at default. I'm going to keep everything at default--VPC and subnet.
* As far as auto-assign IP address, again, I'm going to leave it at default here. But from the drop-down menu, you can enable or disable it if you choose to do that.
* I'm not going to assign anything from an IAM perspective, and I'm not going to do anything from a monitoring perspective either.
* Under Advanced, I'm also not going to do any kind of advanced settings here. I just want to show you how this kind of works.
* Now, once you're ready, we're going to go to Review.

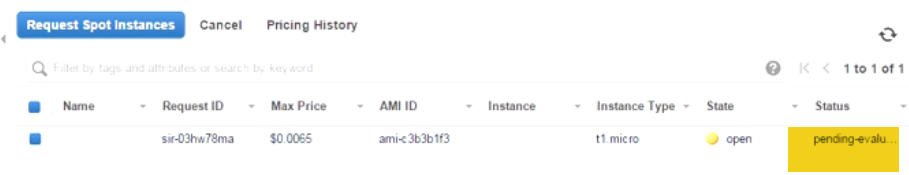
 

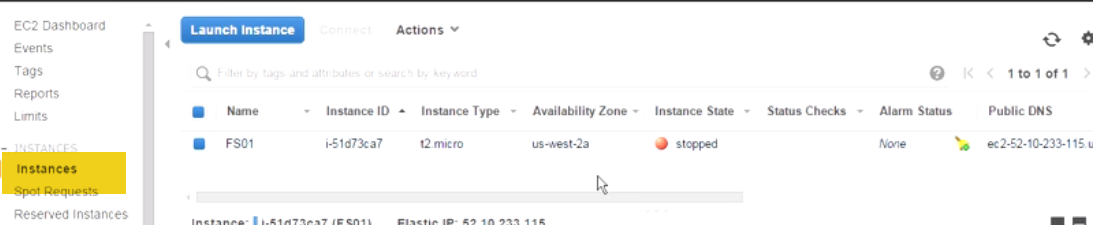
Now you're going to select the key pair here. I already have the WBC key pair. I acknowledge that I have it. And let's 

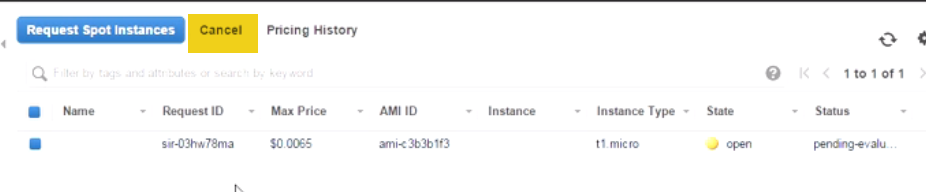
go ahead and request a spot instance. The request has been submitted. You can go here to view the spot instances. Right now it's pending, so the request has not been fulfilled yet. Once the request has been fulfilled, then you're able to go to your instances, and you're able to launch your instances in a regular way.

**Now keep in mind, if someone outbids you, then it's not going to notify you that, Hey, somebody outbid you. It's just going to essentially drop you, and it's going to give that instance to somebody else. So just keep that in mind when you're using spot instances.**

Let's go ahead and refresh this. I don't think this is going to fulfil any time soon here. So this might take a little time, again, to fulfil..

Once this is fulfilled, then if you go to Instances here, 

You’ll see that your instance populates, and then you can launch it and then just run it the same way that we've done with others. 

If you wanted to cancel this particular request, then, again, it's pretty easy. Goto SpotInstances->You can click on the Cancel button up here. You can also right-click and click on Cancel and just come down here it says, Yes, cancel this. 

Now if you cancel it, also remember to go back into Instance and terminate that instance just to be on the safe side. So, again, some things to massage when you're dealing with spot instances.

**Module Summary**

We started off by talking about the instance types. We talked about the fact that you have on-demand, you have reserved instances, and you have spot instances.

We also talked about different characteristics, what you get with each, the hourly rate, how that changes, what are the different pricing models, what you get with RI versus spot, etc.

And then we moved on to move on to talk about EC2 instance families. We talked about the general purpose, the compute optimize, the micro, the memory optimize, the GPU optimize, the storage optimize.