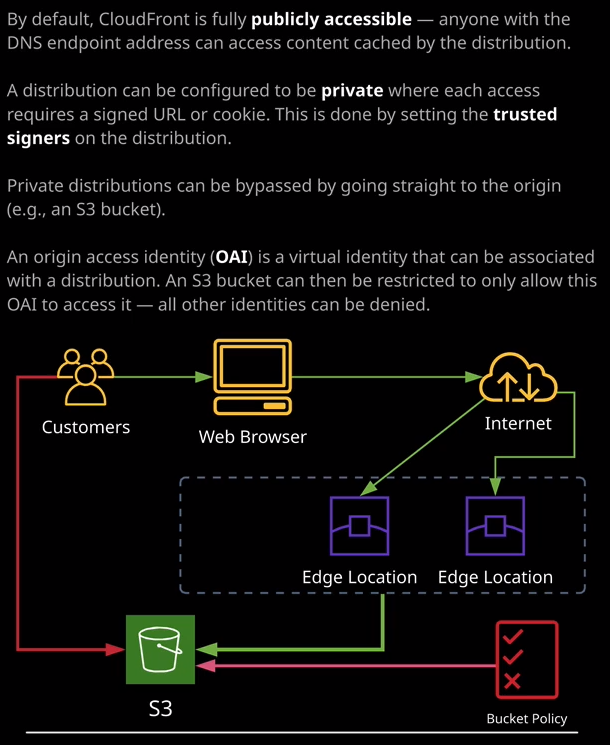
how you can restrict an S3 bucket so that it's only accessible via a CloudFront distribution and that's using a technique known as an **origin access identity or origin access identifier, and that's abbreviated to OAI**.





So if we just look at this diagram down at the bottom my screen for a second, we've got our customers and our customer is using a web browser transiting over the public internet access our edge locations as part of our CloudFront distribution and I want you to imagine for a second that I've configured this distribution to be private, so it's not accessible for anyone in an unauthenticated way. Let's say that it's got some stock images or some sensitive medical data, and our customers are accessing via an application that we've got running inside their web browser. Now, using the architecture that I've discussed to this point in the course, there's nothing to stop the customers finding out directly the address of the S3 bucket and browsing to it directly. So, in effect, bypassing any configuration that we've got inside this CloudFront distribution and accessing all of the data in an unrestricted way. What we want to do is to be able to stop that and that's where origin access identities come in handy. I want you to think of origin access identities as a virtual identity that you can create, associate with one or more distributions, and then use in bucket policies so you can restrict that bucket so that it can only be accessed using the origin access identity. Now, the easiest way to show this is to actually demonstrate it. So in order to do that, first, I'm going to click on "Services" and I'm just going to browse to the S3 console, because what I want to do is open up our global cat website, both using CloudFront and direct S3. So I'm going to select the bucket and I got to go to properties and then static web hosting and then I'm going to copy this endpoint URL into my clipboard and open it in a new tab. I may as well use the same tab, so I paste it in and open it up directly. So you'll see at this point we can access the global cats incorporated website both directly using S3 as well as using the CloudFront distribution. Okay, so step one is to create an origin access identity that's done from the security heading in the CloudFront consoles so I'll click on here and then create an OAI and I'll just give it a comment, restrict global cats inc. So that just gives us an idea about what this is going to be used for and I'll hit "Create." Now that we've created this, this is actually a virtual identity in its own right, and we can use this in various different places. Now, the first thing that I'm going to do is select this CloudFront distribution, go to distribution settings, then I'm going to select origin and origin groups because I want to restrict this for this particular origin. So I'm going to select our only origin inside this distribution and select "Edit," and then I'm going to restrict bucket access. So I'm going to select "Yes." I want to use an existing identity. Now you can create one as part this process, or you can use an existing identity so that you can reuse it across multiple distributions and I'm going to use this identity that I created a second ago, so the restrict global cats incorporated identity. So what this does is associate this identity with this distribution when accessing this particular origin. Now, by itself, this doesn't do anything. All it means is when this distribution is interacting with our S3 bucket origin, it's simply using this identity. But if you recall when I first set this bucket up in the previous lesson, I configured it so that it was publicly accessible from any unauthenticated or anonymous identity. So I'll need to update that but for now, this is fine. What I'm going to do, though, is grant this origin access identity read only access to our bucket. So CloudFront can do this on your behalf. Now, the way this is done is actually by adding another statement to our bucket policy. Essentially, it allows S3 get object actions, but using the principle of this origin access identity. So rather than allowing any identity access to this bucket, it only allows this particular identity. Now I'll show you how that looks in practice. So I'm going to update the bucket policy automatically as part of this process and click on "Yes, Edit." I'm going to go back to distributions just to see if this is going to take some time and it is it's going to update this. So it will take a number of minutes to deploy this change out to all the edge locations, but what I can do in the meantime, let's click on "Services," open up the S3 control. I'll go to the ac-globalcats S3 bucket, click on "Permissions," and then go to bucket policy and straight away, you'll note that I've now got two statements inside this bucket policy. The first one is the one that I created initially in the previous lesson. This is the one that grants public read access to this bucket to any identity. See, at this point, we got this principle and its a wild card, so any identity is allowed access to this S3 bucket. But now we've got this additional statement which explicitly grants the same level of permissions to this specific origin access identity. So if I left this as is it would not change anything because this statement would apply to any user and this statement would apply specifically to this origin access identifier. What I want to do is delete this top statement, so I'm removing the ability for anyone to access this S3 bucket and only allowing this origin access identity. So I'm going to make that change, then I'm going to click on "Save." Okay, so now that change is made if I refresh using the CloudFront distribution, I've still got access to the content because I'm going via the CloudFront distribution so it's using this identity. If though I go directly to the S3 bucket and hit refresh I'm going to get a forbidden error because now I don't have permissions to go to this bucket directly. Only the origin access identity has permissions now, so I can only access it via the CloudFront distribution.

Now from an exam perspective, **what's important is that you understand the concept of an origin access identity. So you know that it's a virtual identity and you can associate it with a CloudFront distribution and then you can restrict access on an S3 bucket based on the origin access identity. The important thing to understand though is that it does only work for an S3 bucket. You won't be able to use this to restrict access to an EC2 instance or an on-premise web server is an AWS only identity. The other important point that's important to understand the exam is to understand why you'd restrict origin so that it can only be accessed via CloudFront. Now, one of the reasons is where you want to insist on a certain level of user experience. You don't want people going direct to your buckets because they will have a lower level of performance. So that's one reason why you might restrict a bucket using an origin access identity. Another reason is if you do have an application that generates signed URLs to access restricted content using CloudFront you don't want your customers having the ability to bypass that and go in directly to this S3 origin. So that's how you'd restrict an S3 bucket so that it can only be accessed from a CloudFront distribution, and that's using origin access identity or an OAI.**