AWS

Introduction to AWS for Non-Engineers

How did we get in the cloud?

- You've probably heard of the cloud in the past few years referring to ambiguous things that no one quite seems to be able to define. You might have also heard about Amazon Web Services. Perhaps your company is considering utilizing it or you're looking to find out more about this cloud computing platform that's taking the world by storm for your own career advancement. Whatever the reason may be that got you to click on this course, I'm glad you're here. I want to help you start from what even is the cloud to getting excited about Amazon Web Services, cloud computing, and potentially even considering taking the AWS Certified Cloud Practitioner Exam. My mission is to introduce cloud computing and Amazon Web Services to people with non-traditional technical backgrounds. In Introduction to AWS for Non-Engineers One, Cloud Concepts, we will be starting from the beginning, and by beginning I mean the 50s, to begin exploring how cloud computing came to be, why it's important, and how Amazon Web Services, or AWS, fits into the picture. We will also be reviewing various cloud computing concepts that will help you begin studying for the AWS Certified Cloud Practitioner Exam, which is the most fundamental certification exam that AWS offers. I can't wait to begin our cloud journey. Let's get started.

AWS Certified Cloud Practitioner exam

The AWS Certified Cloud Practitioner exam confirms a candidate's overall understanding of cloud computing and Amazon Web Services. It also provides an industry-recognized validation of their knowledge. Candidates are encouraged to have at least six month of experience with AWS Cloud, as well as basic understanding of IT services and their uses within the AWS Cloud infrastructure. The exam is 100 US dollars, and candidates have 90 minutes to complete the multiple choice exam at a test center or remotely at home. Upon completion of the exam, they immediately get a pass or fail notification. The exam is available in English, Japanese, Korean, and simplified Chinese. The Certified Cloud Practitioner exam covers four domains which are, cloud concepts, security, technology, and billing and pricing. The second largest portion of the exam is the cloud concepts domain, which makes up 28% of the exam. You would be asked to define the AWS Cloud, why it's desirable over alternatives, identify aspects of AWS Cloud economics, and describe the different cloud architecture design principles. The security domain makes up 24% of the questions on this exam. This section asks you to define the AWS shared responsibility model and the AWS Cloud security and compliance concepts. It also wants you to identify AWS access management capabilities and ways to find resources for troubleshooting security related issues. The largest portion of the exam is the technology domain, which makes up 36% of the exam. You would need to define methods for deploying and operating IT applications in the AWS Cloud, define the AWS global infrastructure, identify core AWS services, and identify ways to contact or receive technical support if you ran into issues. Last but not least, the billing and pricing domain makes up 12% of the exam. While this domain makes up the smallest portion of the exam, the questions are a little trickier because you need to memorize the different pricing models for AWS. They also want you to recognize the various account structures in relation to AWS billing and pricing, and identify resources available for billing support. The AWS Cloud Practitioner certification exam wants you to have a very broad understanding of AWS Cloud, but you do not need to do anything hands-on in order to pass the exam. All of the questions are multiple choice and completed at the test center or on your laptop at home. **While working within AWS Cloud is definitely recommended, it is not mandatory for you to have six months of hands-on experience before you have enough knowledge to pass the exam.**

Cloud Concepts domain

There are four domains in the AWS Certified Cloud Practitioner Exam. They are cloud concepts, security, technology, and billing and pricing. The four courses in the Introduction to AWS for Non-Engineers series follow these four domains. This first course mainly covers the cloud concepts domain of the certification exam and provides vital fundamental information about cloud computing, its history, and the major players in the cloud computing platforms. Cloud concepts is the second largest domain in the exam and goes over many of the core concepts and benefits of utilizing cloud computing services. For this portion of the exam, AWS wants you to define the AWS Cloud and its value proposition, identify aspects of AWS Cloud economics, and list the different cloud architecture design principles. In this course, you will contemplate questions like what exactly is cloud computing? What makes it different from legacy IT infrastructure? How is the way you pay for resources different from buying and setting up the hardware in your on-premises data centers? You will also learn about concepts like well-architected framework for building IT infrastructure, types of cloud computing, type of cloud computing deployments, and advantages of cloud computing over legacy IT infrastructure. Let's get started with learning about what the cloud is, why cloud computing is taking the world by storm, and what Amazon Web Services, the largest cloud computing platform in the world, is used for.

What is the cloud?

The cloud, you've heard it, it's on the cloud, we can do that on the cloud, back it up to the cloud. It's impossible to escape it in the tech world, but what exactly is it? I imagine you have a big idea of what the cloud is as something floating up there in the sky that has something to do with data storage, somehow. Yeah, you're on the right track. In reality the cloud is actually just a new hip way to refer to the internet. Nothing mysterious at all. We all use it every day. But now we have to consider, do we really know what the internet is? In the most fundamental sense, the internet is a worldwide network of billions and billions of devices. These devices can be computers, servers, cell phones, tablets, or Amazon Alexas. As long as they are connected using the global network, any computer can communicate with any other computer thanks to the internet. And despite things like wifi and cell phone services making it seem like the internet is up there in the sky, the internet is still created and connected using physical cables. These can be TV cables, fiber optic cables, or copper telephone wires. When you open your favorite web browser and type in a website URL, it sends a request. The request is sent to your internet service provider, which is the company you contract with to provide you internet service. Your internet service provider then sends your request to a server which searches for the domain name you requested. If it finds a match it will route your request to the IP address of the server hosting the website. Once your request hits the server of the website you wanted to load, the server responds by sending the web page in little packets back to your computer. The packets are very small and acts like jigsaw pieces that your computer reassembles to load the whole entire web page. So the cloud is the internet and the internet is a global network of billions and billions of devices communicating with each other. And you, as a user of the internet, and the cloud, facilitate the transfer of information across the world faster than ever before.

What is cloud computing?

So, we have this cloud that connects all of our devices together. Through it, we can pass information back and forth, store data, and do other cool stuff. You have to be wondering, what powers the cloud? How does it get its energy? How does it stay on? The answer to that, my friends, is cloud computing. You're used to storing files on your hard drive, right? Sometimes, it's a pain to transport that data wherever you go. I know I've forgotten that precious USB stick before. Well, cloud computing solves that issue. It lets you access your data from wherever, as long as you have an internet connection. The technical definition of cloud computing is the on-demand delivery of compute, database storage, application, and other IT resources. This means through cloud computing services, you have instantaneous access to computational, storage, and software using the internet. Computing resources available when you want it, where you want it. So, cloud computing is quite convenient. It's also very flexible when it comes to cost. Instead of the traditional buy first to use model, like when you buy a computer or car, cloud computing utilizes the pay as you go model. This means that you only pay for resources you use when you use them. For example, say you wanted a server to run your applications on. In the traditional way, you would have to go through a procurement process at your work to find an appropriate server with all the necessary bells and whistles. You would then have to make sure the capacity you are purchasing for isn't too much or too little. Then you have to get the quote from the manufacturer and then wrestle with the finance department to get the budget approved and device purchased. If the demands from the applications are much higher or lower than expected, you have to go back and go through the procurement process all over again to get a more appropriate server. Your department or company needs to have the funds to then purchase that equipment outright. Cloud computing allows you to pay to use only as much server space and capacity as you need at that moment. When you need more or less, you can adjust the rented capacity and your monthly bill will adjust along with it. Instead of a large overhead bill on purchasing a piece of hardware that may or may not even match your needs, you get a monthly statement billing you only for as much as you used last month. Cloud computing facilitates collaborations by allowing you to hold virtual meetings, edit documents together, and communicate via email or messaging services. Many of these services that used to cost a lot of money to purchase and maintain can now be purchased by bootstrap startups for $10 a user a month. Cloud computing allows for instantaneous access to computational, storage, and software resources using the internet when and where you want it. It allows for increased flexibility and affordability, because you are only charged for what you consume when you consume. They allow what used to be only possible with big corporate IT budgets to almost anyone with internet access and a few dollars.

A brief history of the cloud

How did the idea of cloud computing develop? To start from the beginning we have to go all the way back to the 1950s with the invention of mainframe computing. Mainframe computing is the concept of having a central computer accessed by numerous user devices. The central computer which had all the compute capabilities was called the mainframe computer. All the user devices which sent requests out to the mainframe computer were called dumb terminals. These days if you peek into a college computer lab there are computers at every desk fully independent from the ones around it. Back in the '50s however, computers were extremely expensive to buy and maintain. So instead of placing one at every seat organizations would buy one mainframe computer and allow the dumb terminals to share its compute resources. In the '70s the concept of virtual machines emerged. Virtual machines are multiple complete operating systems that quote unquote live in a single piece of hardware. For example, you can have multiple Windows virtual machines living in your single Mac laptop. Suddenly, one single mainframe computer could have multiple operating systems running at the same time to do many different things. Then, a new idea hit them. What if we could use lots of mainframe computers' resources as if it's just one computer? This was the beginning of the modern concept of cloud computing. To make pooling resources a reality developers created a software called a hypervisor that could be installed onto multiple pieces of hardware such as a server. They could then link all of those hardwares and use their combined computational storage powers as one giant resource. Imagine the amount of storage and computing power you can harness by adding up all the memory and hard drive space of every computer in your office. Programs will run super fast and you can store a lot of files and you will be able to analyze data at blazing speed. This is what cloud computing allows people to do in an extremely large scale using the Internet to connect end users to huge computational hardwares in their data centers.

Cloud computing in daily life

The cloud, it's everywhere. I bet you use it more than you think. Let's start with email, something you use everyday. Cloud computing powers the storage and transfer of your messages so you can communicate with your friends and family. Streaming services. You know, Netflix, Hulu, that kind of thing. When we stream movies and TV shows, we benefit from streaming of video resources using cloud computing. These companies store video files on a cloud computing platform and allow thousands of people to access the same video at once. The bandwidth necessary for each video changes every second and cloud computing services can adjust it depending on the traffic. Many people use the personal account features of Office 365, like Outlook.com, without realizing it's a cloud computing platform. Google Cloud encompasses many of the features of your Google account, which includes popular services like Gmail, Google Drive, Google Photos, Google Hangouts, Google Calendar, and YouTube. Google Drive is a cloud storage platform for your photos, files, and videos. You can set up automatic backup to the Google Drive so you can save important files directly from your computer's hard disk to the cloud. It can also be used a cloud computing power collaboration tool for you to edit files with your colleagues or friends. You can get real-time feedback from your collaborators and as soon as a change is made by another user, you can see it reflected on your browser. For many popular cloud computing services, there is a free tier that most of us use. If you decide you need more space or features, they have subscription plans available for all different levels of computational needs, such as more space or accounts. The change in the amount or type of resources you can access is instantaneous and you can immediately enjoy your expanded storage space or added services. When you no longer want the extra perks, you can change your account tier again and from then on you will only be charged for the new amount of resources you're consuming. Cloud computing services allow us to consume, produce, and collaborate like never before and instantaneously. Many of them are quite inexpensive and some are even free. Can you think of other popular websites or services you often use that utilize cloud computing?

Wrapping up: Cloud computing

The cloud and cloud computing permeate many parts of our daily lives as we utilize the internet for work, school, and personal life. We check emails, post on social media, share documents via online file sharing services, stream hours and hours of videos, and use cloud-based car navigation apps. In this wrap-up study break, we will be reviewing concepts like what is the cloud, what is cloud computing, quick history of the cloud, and why we utilize cloud computing over legacy IT infrastructure. Let's get started. When someone mentions the cloud, they are referring to the internet. The internet is made of copper wires in a global network of billions and billions of devices. These devices can be computers, tablets, cell phones, Google Home, really anything that can connect to the network to send or receive information. Even though most people did not begin using the internet on a daily basis until the late '90s, the concept of cloud computing dates all the way back to the '50s, with the development of mainframe computers accessed by dumb terminals. Dumb terminals themselves didn't have any compute powers, but users could send queries to the mainframe computers using the terminals. Development of virtualization together with a piece of software called the hypervisor allowed us to think big. And by big, I mean pulling together multiple servers and using all of their compute and storage resources together as if we're using one extremely large server. While in the past, the amount of resources you could link together was limited by what was in your physical data center, with cloud computing, you have the ability to access as much resources as the service provider can give you. With the internet, you have almost limitless potential with as much computing power as you can get. It's almost as though we're back in the '50s with our laptops and desktops serving as the dumb terminals, and our cloud computing service providers hosting the mainframe computers. Instead of connecting to the mainframe computer in the data center on the same floor, we use the internet to connect to the countless servers linked by hypervisors through big service providers like Amazon Web Services, Microsoft Azure, and Google Cloud.

Advantages of cloud computing

So, why cloud computing? Think of it this way, you don't have to buy a computer with a huge hard disk because you can save your files on the cloud using services like Dropbox. Not only that, but these files are available from any machine connected to the internet. You don't need to buy an expensive gaming computer to play graphics-heavy games because you can use a web service like Parsec to play games on their server loaded with enough memory for seamless playing. Instead of paying hundreds of dollars for a program that you install onto your computer, you could pay a $10 a month, subscribe to a service, to use it through your web browser. How does this work for large businesses? If you are thinking in terms of your company's IT infrastructure, you no longer need to have people setting up physical servers and cabling. The countless hours and dollars spent maintaining a server room and the technology inside can now be used elsewhere. Even things like temperature regulation in a server room could be a source of headaches when setting up the IT department. When a piece of hardware breaks, you have to go through the whole procurement and setup process, which could take months. Cloud computing makes all of that the service provider's issue to solve and gives you a set fee to use the services over the internet. With cloud computing, you pay only when and what you consume. This avoids the overhead cost of buying too much physical space on-premises. If you are about to run out, you can simply scale up on the cloud within minutes. I don't know about you, but I'm into saving money wherever possible. The cloud computing service providers take care of the physical infrastructure and their own huge data centers, so you and your engineers can worry about other more interesting things, saving you manpower and money. You also benefit from the massive economy of scale, since larger cloud computing providers buy their capacities in huge quantities, they are able to offer a portion of their capacity to you for a much lower price than if you try to go out and get it for yourself. There are six major advantages to cloud computing. You can trade capital expense for variable expense. Benefit from massive economies of scale. Stop guessing about capacity. Increase speed and agility. Stop spending money running data centers. And go global in minutes. You no longer have to worry about buying too little or too much of something, and you only pay as you go, allowing you to focus your attention and money elsewhere.

Cloud computing models

There are three main cloud computing models, software as a service or SaaS, platform as a service or PaaS, and infrastructure as a service or IaaS. Infrastructure as a service also known as IaaS refers to the basic building blocks of cloud IT infrastructure. You have control over the networking, security, computer, and servers. IaaS provides the most flexibility and management control of all the different types of cloud computing models, and it's the closest in features to having the traditional on-premises data center. Some examples of infrastructure as a service platforms are Amazon Web Services, Microsoft Azure, and Google Cloud. You can modify and control almost all parts of the infrastructure in the cloud to fit your needs without having to purchase or manage actual hardware. Platform as a service known as PaaS allows you to deploy and manage applications without worrying about the underlying hardware infrastructure. Services offered could be web servers, databases, operating systems, or environments where you can execute specific programming languages to host applications. Some examples of PaaS are Microsoft Azure web hosting, Google App Engine, and Heroku. You can focus on deploying the applications instead of the operational side of deployment. PaaS is different from IaaS in that there is less flexibility as packages are preconstructed. But, you also have to deal with less of the infrastructure deployment and maintenance allowing you more time and resources to focus on the project at hand as opposed to the infrastructure. Software as a service, SaaS, describes completed products managed by the service provider. You get the whole package of the service complete with user interfaces. It's ready for use by an end user regardless of their technical backgrounds. You don't have to worry about how the service infrastructure is maintained or managed. You only have to worry about how you might use the service to fulfill a need. A very popular example of SaaS platform is a cloud-based email service such as Outlook and Gmail. As the user you only need to create an account to log in to send and receive emails, no need to worry about anything else. In terms of complexity and level of involvement required, infrastructure as a service is the most involved and requires the highest level of technical knowledge to execute, followed by platform as a service. Software as a service generally does not require much technical knowledge and it's extremely intuitive and features are ready to use quote unquote out of the box. So you and your team will have to decide which choice is right for your needs and circumstances.

Cloud computing deployments

Analogous to different cloud computing services, there are also different deployment models that have organizations deploy their cloud infrastructure. And thankfully, their names are fairly intuitive. Cloud, on-premises, and hybrid deployment. When an organization utilizes cloud deployment, it means that all parts of its IT infrastructure reside and run on the cloud. All applications were either migrated to or created in the cloud. And the organization relies on internet and their cloud-computing service providers to fulfill their computational and IT requirements. Many small startups utilize this model, as it allows them to be flexible and scalable in their resources while removing the roadblock of costly and time-consuming procurement and management processes for on-premises infrastructure. They may use services like Office 365 for emails, Microsoft Teams for on-demand communication, and Microsoft Azure for their app development and hosting. All resources in a cloud deployment infrastructure live on the cloud. With on-premises deployment, often referred to as private cloud, organizations use virtualization to deploy resources in their on-premises data centers. In many cases, the execution of on-premises deployment looks like the traditional IT infrastructure with its servers, network cables, and data center management. The setup does not provide a lot of benefits of cloud computing. The resources are not accessed using the internet because they are on-site. This means you can access them really quickly because nothing has to be uploaded or downloaded using the internet. However, it could utilize application management and virtualization technologies to increase efficiency of the available resources, such as by deploying virtual machines and internet resources behind a firewall. On-premises deployment provides dedicated resources which means that the organization is not sharing any part of their resources with another organization. This may be a requirement for certain industries that take data privacy very seriously, such as the medical field. The last type of deployment is hybrid deployment which connects on-premises tech with cloud-based resources. This is a very common setup for many established companies that already have their own on-premises data centers, but are in the process of migrating over to the cloud. Hybrid deployment allows organizations to extend and scale their infrastructure into the cloud while still maintaining access to on-premises resources living on on-site servers. Another common use case is to use the cloud deployment as backup in disaster recovery solution. An organization can maintain a working copy on premises, but make sure they have durable backup in the cloud. Because migration of existing IT systems take a long time and is costly, hybrid deployment is a very effective in-between as resources are migrated to the cloud. Flexibility, scalability, and finding your perfect fit are features of cloud computing that shine when considering which model of cloud computing deployment is the best fit for your organization. For organizations that don't have very many IT resources deployed yet, cloud deployment would allow them to utilize the complete flexibility and affordability which are signatures of cloud computing services. For those who need all of their data secured and on-premises, either due to retrieval speed or security requirements, private cloud utilizing virtualization of legacy resources is a good fit. For companies with legacy IT resources that would take a long time to upload to the cloud, but would like to extend their computing stores capacity economically, hybrid cloud deployment might be preferable. Many companies utilize hybrid cloud deployment to have quick access to on-premises resources, but have a very safe backup in case of an emergency.

Design principles of cloud computing

Knowing how to create a well-architected infrastructure allows organizations to build the most secure, durable, efficient, and high-performing IT infrastructure possible. So, how can we do that? First, avoid unnecessary costs. Use only what you need and turn off any servers or resources you aren't using. Reserve resources in advance if you know you'll need a certain amount of compute power, as many services give discounts for reservations and upfront payment. Don't forget to continue monitoring for more ways to optimize as your organizational needs change, and know which resources are causing which charges on your bills. Best practice number two, reliability. A reliable system has the ability to recover from service disruptions often by itself. They can also dynamically adjust computing resources to meet demand. You should be testing your disaster recovery settings and incorporating redundancies in your infrastructure. Redundancy refers to the concept of having duplicate copies of resources so that when one goes down, the other can take over to provide seamless cut over experience for end users. The third best practice, efficiency. Performance efficiency is the ability to use computing resources to adjust to system requirements. It should allow for more experimentation and when a change is set in motion, should be able to go global in minutes. A fourth best practice to consider, infrastructure security. This includes security of information, systems, and assets. Security best practices should be automated. Data should be protected in transit and at rest, which means when it's being moved from one location to another, as well as when it's being stored. For example, when someone sends an email, the data is in transit, being transferred using the internet. If you have a file uploaded to a server, it's at rest. Traceability should be enabled, along with strong identity foundation. This means that in case of a security breach, you are able to see who did what at any point, because every user has a unique user account or access key. Who can do and access what should be well defined and followed. The fifth best practice is operational excellence. This is your ability to run and monitor systems while constantly improving processes and procedures. Everything should be documented and operational procedures should be frequently refined. Failures should be anticipated and learned from, and systems and processes updated to take them into account. If an incident occurs, such as a major service downtime, the whole team should come together to discuss what went wrong, how it could have been prevented, and set up procedures in case it happens again. So, there are five best practices when architecting cloud-based IT infrastructure, cost optimization, reliability, performance efficiency, security, and operational excellence. When all five pillars are taken into account and optimized, you will have a highly performing, stable IT infrastructure that allows your organizations to save money, time, and resources.

Study break: Reviewing cloud computing

Welcome to the Study Break for cloud computing concepts that will come up in the AWS Certified Cloud Practitioner Exam. The major concepts to remember are, the advantages of cloud computing over legacy on-premises IT infrastructure, cloud computing models, types of cloud computing deployments, and design principles of cloud computing, such as the as the Well-Architected Framework of a solid cloud computing IT infrastructure. Let's start with the advantages of cloud computing over legacy on-premises IT infrastructure. AWS calls these the six advantages of cloud computing. The advantages are, trade capital expense for variable expense, benefit from massive economies of scale, stop guessing about capacity, increase speed and agility, stop spending money running and maintaining data centers, and go global in minutes. There are three cloud computing models and there are three cloud computing deployments. The cloud computing models are, Software as a Service, SaaS, Infrastructure as a Service, IaaS, and Platform as a Service, PaaS. The cloud computing deployments are, public cloud, hybrid cloud, and private cloud, otherwise known as on-premises cloud. Finally, the Well-Architected Framework of cloud computing provides best practices framework for designing a stable, robust, and secure IT infrastructure on the cloud. The five pillars of a Well-Architected Framework are, cost optimization, reliability, operational excellence, performance efficiency, and security. If you are unsure about any of the concepts mentioned in this video, feel free to pause and go back to the specific videos. Knowing these concepts and models could mean a few extra points on the exam which could go a long way in securing you the certification. Most importantly, the six advantages of cloud computing comes up again and again in the exam, so it's well worth your time to make sure you know what these phrases mean in layman's terms.

A brief history of AWS

Add to cart, checkout, confirm payment. Rejoice. I imagine most of you have experienced this cycle, that visceral joy you get from online retail therapy, and no other company is better at facilitating it than the e-commerce giant, Amazon. A company had a market cap of $1 trillion in 2018 becoming only the second company in the United States to ever hit that mark. Amazon was founded by Jeff Bezos in 1994 as a humble online bookstore before most of us even considered buying anything online. Amazon

Web Services, or AWS, the cloud services platform did not come around for almost a decade after the bookstore turned retail giant was founded. And it brought a completely new side to their business. The framework for Amazon Web Services was launched internally within Amazon all the way back in 2002. At that time, it was called Amazon.com Web Service. Amazon was planning to launch merchant.com, an e-commerce service that helped third-party shops create online shopping websites using Amazon's e-commerce engine. Developing this platform helped to pave the way for Amazon to evolve from an online store to a service company. Surprisingly, it took several years for any real competitors to arrive in the cloud computing platform arena, which has contributed to AWS maintaining its majority market share in the industry. However, the gap is quickly being bridged by other large cloud computing platforms like Microsoft Azure and Google Cloud gaining more and more of the market share every year. It has only been a little over a decade since the very first service was launched in AWS, but the platform has grown exponentially both in its customer base and service offerings. The cloud computing platform currently has over one million active customers. And in 2017, 10% of Amazon's revenue came from AWS.

What is AWS?

Before I even knew what the cloud meant or what cloud computing does, I had heard of AWS and knew that it was a big deal in the tech industry. It took only a quick Google search to find out that AWS stands for Amazon Web Services. From the name, I surmised that it was an Amazon product. I was however, completely unprepared for the extent of the types and number of services AWS provided for organizations all over the world. AWS offers IT infrastructure services to businesses and organizations as web services to help them scale and grow efficiently. AWS provides what used to be purchased as hardware, such as network switches and servers, as resources to be accessed using the internet. Because of cloud computing's pay as you go model and robust resources, organizations are able to save time, money, and human resources by moving their resources to AWS. As of winter 2020, there are 24 groups of services offered by the platform ranging from compute to storage to game development. Each group contains anywhere between one to 12 services with more being added all the time. You can host your static files using simple stores service, host a WordPress blog using elastic compute cloud, send emails using work mail, stream desktops using workspaces or create games using game lift. It's probably not far from the truth to say that your imagination is the limit for what you can potentially architect and create using Amazon Web Services. As a cloud computing service provider, AWS boasts flexibility, scalability, and reliability alongside affordability that was impossible with traditional on premises IT infrastructure. With AWS, engineers can concentrate on building your products and features instead of worrying about the it infrastructure's ability to handle their scaling.

Big companies using AWS

So you get it. Amazon Web Services is a pretty big deal and a lot of companies use the platform to serve their computational, storage, hosting, and IT infrastructure needs. AWS has more than a million active customers ranging from Airbnb to General Electric. Let's see how various companies are using AWS to power their infrastructure, to work for them in the background, so they can focus more on growing their businesses. Utilizing the hybrid cloud deployment model, Comcast built an app for Xfinity services that links AWS Cloud and their on-premises data centers seamlessly. Comcast is the world's largest cable company, and thanks to having their hybrid environment, they're able to deploy features to Xfinity X1 several times a week instead of every 12 to 18 months, which was the timeline with their old architecture. Expedia, your friendly travel companion, is in the process of migrating 80% of its mission critical apps to AWS within the next few years. Expedia provides travel related services through expedia.com, and 200 other travel booking sites all around the world. Because of their extensive global footprint that requires continuous updates and innovations, they chose AWS to host a new service called Expedia Suggest Service. At the time, AWS was the only cloud service provider that supported the Asia-Pacific customers. This made them a great fit as the global travel company serves customers from all over the world, including Asia. Investors use Dow Jones to learn about the going ons of the financial markets around the world. When the lease of their physical data center in Asia hosting the Wall Street Journal for agent customers ran out in 2013, they moved to AWS. Now, all of their Asia traffic is running through AWS, and the transition has saved Dow Jones 25% every year over the cost of leasing a data center. Atlassian, who owns popular product and project management tools, such as JIRA and Confluence, uses AWS to scale and enhance availability and disaster recovery. Breakfast cereal tycoon, Kellogg's, has tight margins, and estimates that it will save a million dollars over the next five years in software, hardware, and maintenance costs, by using AWS. Some other companies you might have heard of running on AWS, are Citrix, Square Enix, Spotify, USDA, UK Ministry of Justice, and Netflix. Anyone choosing to run their cloud infrastructure on AWS, will be in good company.

Popular services offered in AWS

Compute services provide virtual server hosting, container management, and serverless computing. You can set code to run to certain triggers using Lambda, run virtual machines using Elastic Compute Cloud or EC2, quickly set up and run small websites using LightSail, or create a unit of software to ship out to your users using Elastic Container Services or ECS. Compute services are backbones of cloud computing platforms as they provide the much-coveted computing resources that many companies are looking for. Instead of having to host their own servers in their own data centers, they can rent servers from AWS for pennies on the dollar. Storage services provide storage for both in-use and archival files. You can use Elastic File System or EFS to create shared folders in the cloud. You can upload flat files like images, videos, or text files to Simple Storage Service or S3 and link to it directly to use on your website. You can also archive files and store large amounts of data for cheap using Glacier or you can use Storage Gateway to take daily backups of your company's on-premises data and send them to the cloud for safekeeping. Storage solutions are cheaper than ever with cloud computing and AWS provides many options depending on the frequency of access and durability of data you require. AWS also offers fully-managed relational and NoSQL databases. Their cost-efficient relational database is called Relational Database Service or RDS and a highly scalable NoSQL database is called DynamoDB. They also offer a fully managed, easily scalable petabyte-scale data warehouse service called Redshift and a highly scalable caching service called ElastiCache. ElastiCache allows you to run extremely intensive computations by caching necessary data in the cloud. All of the database services are highly scalable and cost efficient so you can crunch all the numbers and data you need for a fraction of the cost of an on-site database server. It's easy to get lost in the abundance of options but it's also exciting to consider the almost limitless potentials in what we can create using these resources.

Create an AWS account

Let's dig right in and create an AWS account. You'll need to go to aws.amazon.com. Just as a warning, creating an account requires to have a valid phone number and a credit card. To create the account, click on the very aptly labeled button here that says "Create an AWS Account." Fill in your email address and create a password. An AWS account name is a unique username for AWS. You might have to try a few times before you hit an account no one has taken yet. Once you're done filling out the form, click Continue. For the account type, choose Personal, as you are creating this account to learn and explore. Enter your full name, phone number, as well as your address. Make sure to read the agreement and check the box before proceeding. Once you're done, click Create Account and Continue. We're almost done. The next page asks for your payment information. For the first 12 months after your account creation, you are eligible for what is called the AWS Free Tier. This means that up a certain usage level, you can try out many of AWS's most popular features for free. The payment information is in case you use features that require payment, or if you go past your free tier limits. They will make a small test charge to make sure your payment method is valid. It will go away once your account confirmation is complete. Now click Secure Submit. We're almost done. This page will ask you to verify your account creation with a phone verification. They will call you so you can put in a code to verify that you indeed did create this new account. This number can be any phone number that you can receive calls at, so it can be an extension at work if needs be. Put in your phone number, type in the code in the security check, and click on Contact me. When prompted, enter the four numbers that came up after clicking the button when you received the call. Feel free to pause this video now to receive the phone call. Once you complete the verification, click Continue. Now you'll select a support plan for your account. As you can see on this page, there are three kinds of support plans available to you, which are Basic Plan, Developer Plan, and Business Plan. There is also a fourth one, Enterprise, but that's only for bigger companies that require a lot of support. Each plan has different features, support tiers, and costs associated. You can click on the Basic Plan, as it's free and provides you access to health status and notifications for your various services. You can learn more about each support plan by clicking on Learn More. Now we wait for AWS to finish creating your account. The page you're on now helps you personalize your account by picking your role and interests while you wait for the account to be activated. Once your account is fully ready, which generally only takes a few minutes at most, they will notify you by email. There you have it. You've successfully created an AWS account. Go check your email to log in to your own AWS management console for the first time.

Exploring the AWS dashboard

When you first log in to the AWS Management Console, you'll immediately notice that there's a lot going on. No worries. Most of the time, we only work with one or two of the resources on this dashboard. First off, let's take look at the right top corner of the browser. Here, you'll see your user name you chose for yourself when you signed up for an account. When you click on your user name, you can find out information and have a quick way of accessing your security credentials. The billing portion is important as you start exploring different services, as some of them will cost money, even during the free tier. Let's move on to the link near the user name labeled Support. Here, you can explore different ways of finding documentation and support resources for your issues or questions. You can create tickets in the support center or ask peers in the forum. You can also find documentation and tutorials on how to troubleshoot or create certain functions in the documentations and trainings offered by AWS for free. The Services link at the top left corner of the browser takes you to the list of all the services AWS offers. This list expands as new services are announced, and just in October of 2018, AWS added new categories, like Blockchain and Satellite. Some of the most popular categories of services are Storage, Commute, and Database. You can click on any of them and they will provide you information and introduction to the services, as well as resources you could check out to learn more about them. Back on the main dashboard, there are resources to learn about various services and what they can do for you, such as the Build a solution and Learn to build section. You can also go to the Explore AWS column to the right and see what AWS things you should check out now. There are many parts to the AWS Management Console dashboard, but we've gone over many of the main resources available for you to begin your dive in to AWS. A big portion of learning a new system or technology is knowing where to look for answers. And the support resources available on the dashboard can answer many of your potential questions and issues. Go ahead and take a little while to explore the dashboard on your own.

AWS Free Tier

Go to AWS.Amazon.com/free. AWS Free Tier allows new potential customers to test out and use many services offered by AWS for free. This allows you to become comfortable with many of the services and AWS gains a potential customer. When the 12 months are over, you begin to be charged for services you consume at regular rates. As a warning, you will get a notification when your 12 months are expiring, but you will then need to manually turn down or delete your services if you don't want to be charged. There are three different types of Free Tier offerings. They are 12 months free, always free, and trials. Let's take a look at each option. The first option is 12 months free. As the naming suggests, these are offers that expire 12 months after you sign up for your account. All of the services have usage limits, and if you go above the limit, you will be charged at normal rates even within the first 12 months. Some of the common limitations are use time, number of requests, amount of storage, number of characters, and actions per month. The second option is always free. And, you guessed it, it's always free, up to a certain point. The final type of Free Tier offering is trials. Most of the trials are for less than 12 months, and have stricter limits. Take note, one important thing is that being on the AWS Free Tier plan doesn't mean you have unlimited use of everything. Depending on the service, there are different limits. AWS Free Tier is a generous offering that helps bring in new customers for AWS, and as a brand new user of AWS, it allows you to test out the services and learn and explore this powerful system.

Use case: AWS Free Tier

So, now that we are armed with an AWS account and AWS Free Tier, let's explore a use case for a real project you can create using mostly free resources. Imagine you have to create and host a WordPress website using AWS. You can very quickly spin up an Elastic Compute Cloud or EC2 Instance, that comes loaded with WordPress. AWS has a marketplace for preconfigured servers called Amazon Machine Images or AMI. These are basically templates of servers that you can create and immediately get it preconfigured to a certain way. In this case, a company called Bitnami has created a WordPress AMI called WordPress Certified by Bitnami. It is Free Tier eligible and runs on an Ubuntu server. You will be led through the setup process and once you're through, you will have a WordPress website set up and ready to go. EC2s have fairly long URLs through, which could be something like ec2-52-204-122-132.compute-1.amazonaws.com. That's usually not a very attractive way to introduce your blog to your new friends. You would probably want something like mycoolblog.com to take your visitors to your brand new blog. AWS has a service to help you do just that. The simplest way is to purchase the domain name that you want using AWS's domain name registrar Route 53. A domain name registrar is like a phone book. To visit a website you input a domain name like mycoolblog.com and the DNS finds it in an online directory of IP addresses. It then sends your request to the appropriate server so you can load the website. By purchasing your domain on Route 53 and matching the domain name with the IP address of your EC2 Instance, you can make the address mycoolblog.com load your WordPress website. Now, Route 53 costs a few dollars a year for the domain registration and charges a separate monthly usage fee, however, the monthly usage fee for me is around 50 cents a month and domain registration itself was around $12, so for a whole year of website hosting the costs are fairly minimal. If you were thinking about starting a blog for cheap using AWS might just be the way to go, and it doesn't hurt that you are getting some hands-on experience with different services at AWS. There are many resources available on how to set one up ranging in complexity from a simple one, like what we just did using Route 53 and EC2, to using other services like CloudFront, AWS Certificate Manager and Elastic Load Balancer to help secure the website and make sure it stays up even if someone tried to take your blog down with a DDoS attack. Your creativity can take the reins to create just the project you were dreaming of with AWS Free Tier and other services.

Study break: Exam tips and resources

We began this course with what even is the cloud? Since then, we've come a long way. Let's get prepped for the exam. We've learned about the cloud, cloud computing, Amazon Web Services, and various but essential cloud computing concepts to help begin the preparation process for the AWS Certified Cloud Practitioner Exam. In this video, we will review major concepts you should know about for the AWS Certified Cloud Practitioner Exam's Cloud Concepts domain. These topics are, what is AWS, six advantages of cloud computing, three cloud computing models, three cloud computing deployments, and five pillars of a Well-Architected Framework. We will also throw in some study tips for memorizing certain concepts in preparation for the exam. Let's get started. The AWS Certified Cloud Practitioner Exam is the most fundamental certification exam that AWS offers to help validate the candidate's overall fundamental understanding of the AWS Cloud. It includes four domains which are, Cloud Concepts, Security, Technology, and Billing and Pricing. We began with Cloud Computing and then went over to the Cloud Concepts domain. For the Cloud Concepts domain part of the exam, AWS wants you to define the AWS Cloud and its value proposition, identify aspects of AWS Cloud economics, and list the different cloud architecture design principles. AWS, or Amazon Web Services, is a cloud computing platform created by Amazon and currently holds the world's highest market share in the cloud computing sphere. It provides many different IT services on the cloud and helps to make it easier, faster and cheaper to run your IT infrastructure compared with legacy on-premises IT infrastructure. According to AWS, there are six distinct advantages of utilizing cloud computing over on-premises IT infrastructure. The advantages are, trade capital expense for variable expense, benefit from massive economies of scale, stop guessing about capacity, increase speed and agility, stop spending money running and maintaining data centers, and go global in minutes. Basically, utilizing cloud computing is faster, cheaper, and more agile than utilizing your own data centers. There are three types of cloud computing models and three types of cloud computing deployments. The three types of cloud computing models are, Software as a Service, Saas, Infrastructure as a Service, IaaS, and Platform as a Service, PaaS. The three types of cloud computing deployments are, Public Cloud, Hybrid Cloud, and Private or On-Premises Cloud. Want the way to memorize them? Try out my silly memorization method, SIP PHO. SIP is an acronym for the three cloud computing models, Software as a Service, Infrastructure as a Service, and Platform as a Service. And oh boy do I love myself a good bowl of Pho. PHO stands for the cloud computing deployment models, Public, Hybrid, and On-Premises or Private Cloud. There are five pillars of a Well-Architected Framework. They provide best practices for specific areas of running an AWS Cloud IT infrastructure. To help me memorize these five pillars, I created the acronym CROPS. The pillars are, Cost Optimization, Reliability, Operational Excellence, Performance Efficiency, and Security, CROPS. What do you think? Do you think you'll be able to answer questions in regards to all the topics we learned about in this course? If not, please don't hesitate to go back and rewatch some of the videos, and take notes.

Next steps

- Well, that was a lot of information in such a short amount of time. I'm so glad you stuck with me to the end. I hope you not only learned a few things, but enjoyed the process, too. If you are interested in learning more about Amazon Web Services, and even potentially taking the AWS Certified Cloud Practitioner exam, please check out the rest of the introduction to AWS for non-engineer series here at LinkedIn Learning. The courses cover the four domains of the AWS Certified Cloud Practitioner exam, which are cloud concepts, security, technology, which we refer to as core services, and billing and pricing. If you have questions or want to learn more about cloud computing, and potential careers that work with or in cloud computing, please come visit cloud newbies, a community of cloud newbies and seasoned pros where we learn about cloud computing and study for certifications together. You can visit us at cloudnewbies.com. If you're looking for a resource website while you're beginning your research into Amazon Web Services, you can visit me at awsnewbies.com where I introduce cloud computing in AWS in a jargon-free way. Thanks again for watching and I hope to see you again in one of my other courses or resources. Good luck.

AWS

Introduction to AWS for Non-Engineers

Module 2 Security

Core concepts of cloud computing

- You probably heard of the cloud in the past few years, referring to ambiguous things that no one quite seems to define. You might have also heard of Amazon Web Services. Perhaps your company is considering utilizing it, or you're looking to find out more about this cloud computing platform that's taking the world by storm for your own career advancement. Whatever the reason may be that got you to click on this course, I'm glad you're here. I want to help you start from, What even is the cloud, to getting excited about Amazon Web Services, cloud computing, and potentially even consider taking the AWS Certified Cloud Practitioner exam. In Introduction to AWS for Non-engineers Two: Security, we will be focusing on security in the cloud and how it is similar and different from security in legacy IT infrastructure. We'll learn about security-related services offered through Amazon Web Services, and various security-related concepts that are important to keep in mind as you consider moving your resources onto the cloud. This course is also a vital part of your exam prep if you are thinking about taking the AWS Certified Cloud Practitioner exam. I can't wait to begin our cloud journey. Let's get started.

AWS Certified Cloud Practitioner exam

The AWS Certified Cloud Practitioner Exam is the only foundational level certification exam from Amazon Web Services, and requires no hands-on engineering experiences or prerequisite certifications. It validates the candidate's overall fundamental understanding of the AWS Cloud and basically serves to show your employers that you get what AWS does and how it operates to provide cloud computing services to its customers. While it is in no way required, it is a recommended stepping stone to taking the Associate or Specialty-level certification exams. AWS recommended that a candidate has six months of experience with the AWS Cloud in any role, ranging from technical, managerial, sales, purchasing, or financial before taking the certification exam. **But, from my experience and those of many others, this is not a strict necessity as long as the candidate has the fundamental knowledge necessary for the exam through studying.** Basic understanding of IT services and their uses in the AWS Cloud platform is also recommended. The certification exam is 90 minutes long and can be taken online or at the testing center. It costs 100 U.S. dollars per attempt and is available in English, Japanese, Korean, and simplified Chinese. Unlike many other AWS certification exams, it has a hard pass score of 70% and is a multiple choice exam. So, what does the certification exam validate in terms of your knowledge and skills? The certification shows that you can define the AWS Cloud and the basic global infrastructure, describe basic AWS Cloud architectural principles and value proposition, and describe key services on AWS Cloud as well as common use cases. It also shows that you know the basic security and compliance aspects of AWS Cloud, and can define the billing, account management, and pricing models. The exam proves that you can describe the core characteristics of deploying and operating your IT infrastructure in the AWS Cloud. Finally, it shows that you can identify sources of documentations and technical assistance such as submitting support tickets and reading white papers.

Security domain

There are four domains in the AWS Certified Cloud Practitioner Exam. They are cloud concepts, security, technology, and billing and pricing. The four courses in the Introduction to AWS for Non-Engineers series follow these four domains. This course that you are watching now covers the second domain, the security domain, because securing the cloud by keeping your AWS Cloud infrastructure safe from both internal and external exploits is extremely important. The security domain is well worth your time to investigate.

There are four major points AWS wants you to be comfortable with before taking a Certified Cloud Practitioner Exam. They want you to define the AWS shared responsibility model and AWS Cloud security and compliance concepts. They also want you to identify AWS access management capabilities and also be able to identify resources for receiving security related support. We will review different security related concepts, like the shared responsibility model, security pillar of the well-architected framework, principle of least privilege, and AWS Cloud compliance. Ever wonder what it means to quote/unquote secure the cloud? We will also talk about what it means to secure the cloud, which could be a little different or similar to securing your on-premises IT infrastructure in a data center. In addition, we will review some major security related AWS services, like AWS Identity Access Management or IAM, AWS Web Application Firewall, or WAF, and AWS Trusted Advisor amongst other core security services. We'll provide study breaks and exam study tips so you can begin preparing for the AWS Certified Cloud Practitioner Exam security domain. Ready to get started? Let's go.

Security in the cloud

Security in the Cloud. If you're anything like me, when you think about security for your data and your IT infrastructure, you probably think of the server room in your office, locked with a card key that only the IT department has. Or maybe your company has a data center off-site that you have to drive two hours to get to, to make sure your backups are being saved securely. When you get to the data center, you probably encounter a lot of security personnel and need to submit a lot of credentials to step into the facility and get anywhere near the servers. This image of securing data is quickly being replaced by cloud-based security where you no longer have to keep costly data centers functioning and secured. Instead, you can have a cloud computing service provider manage their own data centers on your behalf so you can focus on other aspects of IT infrastructure management. When you deploy your IT resources into AWS Cloud, you benefit from the global network of data centers and architecture built with security in mind. AWS helps you keep your data safe in their highly secure data centers, and there are safeguards in place to help protect customer privacy. There are dozens of compliance programs embedded into AWS to help you meet your industry's compliance requirements for data security. Securing your data on AWS Cloud allows you to maintain the highest standard of security without having to manage your own data centers, which saves you time and money. It also allows you to scale the size of your business quickly, as AWS is designed to keep data safe no matter how big or small your cloud usage is. Let's learn about a few concepts and frameworks for security in the cloud before discussing specific security-related services that AWS offers.

Shared responsibility model

When utilizing the cloud to house any part of your technical infrastructure, you must first consider the security impacts of moving your resources onto the cloud. Unlike the on-premises data center that is protected by the virtue of being within your physical reach, data centers hosting your cloud resources are in an undisclosed location in data centers managed by AWS. So who's responsible for the security of the data center? The servers? The networks? The data itself? All those EC2 instances that are running computations. Who makes sure their security patches are up to date? Who protects the data from being corrupted? AWS has a model that helps to puzzle these questions out, called the Shared Responsibility Model. As you might expect from the name, the Shared Responsibility Model asserts that the security of cloud functioning infrastructures is a shared responsibility between the customer and AWS. While there are certain parts of the infrastructure that the customers no longer have to worry about, there are still components that are the customer's responsibilities to secure.

In the most basic breakdown, AWS is responsible for the security of the cloud, while you, the customer, are responsible for security in the cloud. Let's see if we can deconstruct what AWS might mean by this. When AWS says that they are responsible for security of the cloud, they mean that AWS is responsible for protecting the infrastructure that runs all of the services offered by AWS Cloud. This includes hardware, software, networking and the data center facilities that run their cloud computing platform. You can think of it like, AWS is responsible for security of the components that make up the AWS Cloud, like the data centers and physical servers. On the other hand, when AWS says that the customers are responsible for security in the cloud, they mean that the customers are responsible for varying levels of security functions, depending on which services they are using. These could be in forms of protecting customer data, platform, application, and identity or access management. Or operating systems of virtual machines, configuring firewalls and data encryption. You can think about it like, we are responsible for the security of things inside the AWS Cloud, like data encryption and patching servers. There are many granular settings and concepts within the Shared Responsibility Model for AWS Cloud, which you can read on their website. But the basic concept that you need to know for the exam when this model is brought up is this, AWS is responsible for security of the cloud, while you, as the customer, are responsible for security in the cloud.

Well-architected framework

- Security in a Well-Architected IT infrastructure. When you're considering architecting how your AWS cloud IT infrastructure will look, you will have to put in a lot of considerations on how to make sure that it is secure from both external and internal threats. AWS has developed the concept called the five pillars of a well architected framework to help you build the most secure, fault resilient, efficient and high performing IT infrastructure possible. The pillars are discussed in detail in the first course of this series titled Introduction to AWS for non engineers, one cloud concepts. Since this course pertains to security, we'll be doing a deeper dive into you guessed it, the security pillar. They are Identity and Access Management or IAM, Detective Controls, Infrastructure Protection, Data Protection and Incident Response. To make sure user access is manage properly, you would want to implement a strong identity foundation. This would entail utilizing the principle of least privilege, which means that you only provide access to what people need to do their jobs, and no more. We will go over this concept in more detail in the next video.

You should enable traceability by monitoring alerts, audit actions and changes to your environment in real time. Security should be applied on all layers instead of just on a single outer layer of your infrastructure. For a virtual server, this could mean that you make sure your infrastructure is secured at the organizational subnet, load balancer virtual machine in the operating system layers. Security best practices should also be automated, so that you can scale more rapidly and cost effectively. If the security methods are automated, You can just replicate that for every new instance or resource you deploy instead of having to manually set them up. Data should be protected At Rest and In Transit data is At Rest when it is stored somewhere like in an S3 bucket. Data is In Transit when it is moving from one place to another, such as when you send an email from your mail server to your friend's mail server. Security mechanisms should be adjusted depending on sensibility of the data. You should also keep people away from the data by eliminating the need for direct access or manual processing of data. In this way, human error and loss or modification of sensitive data can be prevented. Finally, when a security event occurs, you should be prepared to intervene, investigate and deal with the event. And once the issue is resolved, update the incident management process to learn from the security event. Security is a very vital part of running and architecting a well architected framework. You can strive for stability by focusing on protecting the data and resources against security events. And when an event occurs to learn from the event and update Incident Management procedures.

Principle of least privilege

Principle of least privilege. Who can access what? When you start a new job, you get some accounts to log in. It can be your not so new computer with someone else's coffee stains on the keys, or your corporate email account that has fifty emails waiting for you already. Or, it could be your company's shared network drive on the server, where your team and your predecessors have been keeping documents that everyone needs to access. Say you work in the sales departmnet. You should have access only to resources and information that you require to do your job. That could be the client list for your team, or deck templates for slideshows you will now be creating to present to potential clients. Or, even the products you are selling, however, you will not expect and should not have access to resources like pending legal cases being handled by the legal department, the not-yet released product mock-ups being developed by your dev teams, or list of personnel reshuffling that the HR department is contemplating.

This concept of providing access only to resources that a person needs to do his job, and no more, is called the principle of least privilege. The concept is this. The CEO of the company should have access to a lot of the corporate resources. The newly hired sales associate should not. The IT department should have the ability to administer the services, but probably not have access to the sensitive information and files themselves. Every role has a set of access permissions necessary to effectively complete its job, and the individual in the role should have no more or no less than the optimal level of access.

To follow the principle of least privilege in AWS, you provide access to services and resources for your users and other AWS services by using a service called Identity Access Management, or IAM. When you provide users or services access using IAM policies you should start with a minimum set of permissions and grant additional permissions only as necessary. Determine what the user or service needs to be able to do and craft policies to allow them to perform only those specific tasks. For example, a marketing manager might need to access certain marketing-related S3 bucket to upload flat files for the company's website. You may remember that S3 is a file store service offered by AWS, and buckets are like folders inside the service that holds your files. However, he will not need access to the S3 buckets where air logs are being dumped into by an app being developed by the dev team. The IAM access granted to the marketing manager should provide him only the absolute necessary access in the company's AWS cloud infrastructure. Remember to provide only the minimum amount of access a person or service requires and nothing more to keep your AWS cloud infrastructure secured.

AWS Cloud compliance

AWS has many compliance programs available for your review in order to determine whether your industry allows you to store data or use AWS for your business. You can find the AWS Compliance Programs by going to aws.amazon.com/compliance/programs. Let's take a look together. Here, we can see that the AWS Compliance Programs are divided up into regions of the world. There's also a whole entire section on privacy. That's definitely a big one for the data on the cloud. If your organization deals with patient data in the United States, you would need to be cognizant of HIPAA. You can find out how AWS is compliant with HIPAA and how data is secured on AWS. Let's take a look. Let's see, here we go. Laws, regulations, privacy, HIPAA. If you had any questions, the quick FAQ will respond to many of the common questions. There are many compliance programs available for review. And this page allows you to easily find out whether your resources can or cannot be hosted on AWS if you have compliance requirements for your data. Again, you can find the information on aws.amazon.com/compliance/programs.

Study break: Security domain

What are the differences between security in the Cloud and security in an on-premises data center? Security in the Cloud may look a little different, and include some added benefits. Let's review the security domain of the AWS Certified Cloud Practitioner Exam together. Some topics we'll be covering are the security in the Cloud, Shared Responsibility Model, security pillar of the Well-Architected Framework, Principle of Least Privilege, and AWS Cloud Compliance. Let's go.

One of the biggest benefits of utilizing Cloud computing is that you no longer have to purchase equipment and maintain your own data center to run IT resources. Cloud computing providers like AWS manage the data centers so you can focus on other aspects of IT infrastructure management. When you deploy to the AWS Cloud, you benefit from the global network of data centers and architecture built with security in mind. There are dozens of compliance programs embedded into AWS to help you meet your industry's compliance requirements. AWS is designed to keep your data safe, no matter how big or small your Cloud usage is, so you are free to scale your business as quickly as you want.

There are three major concepts that outline AWS' recommended security practices. These are: the Shared Responsibility Model, the Security Pillar of a Well-Architected Framework, and Principle of Least Privilege. In addition, we will review how AWS accounts for compliance requirements for data and resources stored in the Cloud. The first concept addresses the question who is responsible for security? The answer is slightly complex. You, as the consumer, are responsible for security in the Cloud. AWS, as the Cloud computing service provider, is responsible for security of the Cloud. This concept is called the Shared Responsibility Model, and it asserts that the security of the data and resources in the Cloud is a shared responsibility between the Cloud computing service provider and the customer. While the customer no longer has to worry about certain aspects of IT infrastructure, like securing the physical data center or hardware, there are other aspects that they are still responsible for, including patching virtual service regularly, and utilizing proper permission sets so only people who should be accessing certain resources, do access them.

Next, AWS addresses how can you best protect your AWS Cloud infrastructure from both internal and external security threats? AWS has the five pillars of a Well-Architected Framework to help it's customers build the most secure, fault resilient, efficient, and high performing IT infrastructure possible. Within the five pillars, there is the security pillar, which outlines how you can secure your infrastructure adhering to best practices. Security in the Cloud is composed of five areas: Identity and Access Management, Detective Controls, Infrastructure Protection, Data Protection, and Incident Response. Architecting a Well-Architected Framework can go a long way to making your IT infrastructure stable and secure.

Next is Principle of Least Privilege. What resources should you provide access to? The Principle of Least Privilege states that you should only be providing access to resources that an entity requires to do it's job. Every role has a set of access permissions necessary to effectively execute it's job, and the resources and individuals should have no more or no less than the optimal level of access. In AWS, you would make this happen by using a service called Identity and Access Management, or IAM, providing granular access permissions. Providing the minimum amount of access to entities to complete their work is a vital way to keep your IT infrastructure secure. The Principle of Least Privilege coincides with the Shared Responsibility Model, where the customer, you, are responsible for security in the Cloud by making sure access is provided responsibly.

Lastly, here are many AWS Cloud Compliance programs available to help you determine if your industry allows you to store data on AWS. Many industries have compliance requirements for storing your data, such as HIPAA for medical organizations. You can learn more about the various compliance programs AWS offers by visiting aws.amazon.com/compliance. In the security domain of the AWS Certified Cloud Practitioner Exam, AWS wants you to be able to explain what concepts like the Shared Responsibility Model and Principle of Least Privilege may mean in real life scenarios. If you feel like you want or need a refresher for any of the concepts we reviewed in this video, feel free to go back and watch the videos again.

Identity and Access Management (IAM)

When utilizing the cloud to house any part of your technical infrastructure you must first consider the security impacts of moving your resources onto the cloud. Unlike the on-premises data center that is protected by the virtual being within your physical reach, data centers hosting your cloud resources are in an undisclosed data centers managed by AWS. By now you've probably surmised it's not the best idea to provide unrestricted access to your IT resources to everyone in your organization, but is there an easy way to provide extremely granular access permissions to every user in service but at the same time make them easy to manage? Thankfully, AWS has a service called Identity Access Management commonly referred to as IAM which helps you do just that.

Identity and Access Management, or IAM, is a free service provided by AWS that enables you to manage access to services and resources on the AWS cloud. You can create and manage users and groups as well as set permissions to allow or deny access to various resources. The permissions are global which means that the access you set for a user or group will be true for all regions in AWS Cloud. When providing access to users and services you should follow the principal of least privilege. There are a few ways you can set permissions for various services or users to access your AWS resources. You can use IAM to manage users, manage roles and manage federated users. The first way you can set access is by using IAM to manage users. You can create users in IAM and assign them individual security credentials. These users can have very granular permission sets so you can control which operations a user can perform on which specific service. A user could be administrators that need console access to manage the AWS Cloud account, end users who need to access content in the AWS Cloud account or systems that need the ability to programmatically access data in the AWS Cloud account. Programmatic access means that applications are directly accessing resources in the AWS Cloud as opposed to humans doing the same activity. Another way to set access is to manage IAM roles. You can create roles to manage permissions and control what these roles can do in your AWS instance.

An entity assumes a role and can obtain a set of temporary security credentials to make API calls to your AWS resources. This could be used to provide access to a user from another AWS account to your AWS account such as when an organization has separate develop and production environments. The last way to set access is to manage federated users. By enabling identity federation you can allow existing identities in your enterprise to access your AWS cloud instance without having to create an IAM user for each user. You can use any identity management solution that supports SAML 2.0 or use one of AWS's federation samples. You've probably experienced identity federation in action when you sign up for an online service using your Facebook or Gmail account. In a corporate setting you could have your Microsoft Active Directory users have federated access to your AWS cloud instance using identity federation. Some benefits of IAM are enhanced security, granular control, ability to provide temporary credentials, flexible security credential management, ability to leverage external identity systems using federated access and seamlessly integrating various AWS services within the AWS Cloud infrastructure.

Web application firewall (WAF)

Web Application Firewall or WAF. AWS Web Application Firewall, commonly referred to AWS WAF, protects web applications running on AWS Cloud from common web exploits. As the name suggests, it's a firewall service for your web applications. Fancy that. WAF can protect against exploits that could compromise security or availability of your web apps. It can also protect the application from exploits that could force your app to consume excessive resources which in turn could end up costing you a lot of money. With WAF, you only pay for what you use with no upfront commitments. Web Application Firewall improves web traffic visibility, provides cost-effective web application protection, and delivers increased security and protection against web attacks. It is also easy to deploy and maintain, as you can deploy it on Amazon CloudFront as part of your content delivery network solution or via Amazon API Gateway. There is no additional software that you need to deploy and you can reuse the centrally-defined roles across all of your web applications. AWS WAF is an affordable, malleable, and adaptable protection for your web applications running on AWS Cloud that is easy to manage and deploy.

Shield

AWS Shield. Have you ever experienced Distributed Denial of Service or DDoS attacks, you might have tried to access your favorite social network platform, only to find that the website is glitchy, or nothing is loading on your browser? You might Google the website or complain about it on another social media network. And you might find out that the website is experiencing a DDoS attack. A Distributed Denial of Service Attack or DDoS attack is an attempt to make a machine or network resource unavailable temporarily or indefinitely, most often by making excessive repeated access requests to the website using thousands of unique IP addresses. Basically, a hacker or malicious personal organization would overload the server with access requests so that real users can't access the website because it's to busy.

Have you ever felt extremely overwhelmed by multiple requests coming from all different places? Maybe you worked on extremely busy restaurant when a coworker called out or you had multiple project deadlines at the office while some dumpster fire was going on, that you needed to fix, you likely felt overwhelmed, exhausted and unsure how to deal with the workload. And perhaps you mentally shut down unable to process all the different requests. This is what happens to the servers when it's under a DDoS attack.

AWS has a service called AWS Shield, which provides detection and automatic mitigations to minimize the effects of DDoS attack on your applications. AWS Shield helps to minimize application downtime and latency when an attack happens. There are two tiers to AWS Shield in terms of protection and cost. The standard tier is automatically enabled, free to use and protects your web application against majority of common DDoS attacks. When used with CloudFront and Route 53. You can obtain comprehensive availability protection against all known infrastructure attacks. The second tier is called Shield Advanced and it provides 24/7 access to the AWS DDoS response team. It detects and mitigates sophisticated DDoS attacks with near real-time visibility into the events and integrates with AWS WAF. Shield Advanced provides higher level protections, network and transport layer protections and automated application traffic monitoring. Shield Advanced also provides financial protection against DDoS related spikes for charges for EC2, elastic load balancers, CloudFront and Route 53. It is available globally on all CloudFront and Route 53 Edge locations. This means that your web application can be hosted anywhere in the world but still be protected by AWS Shield as long as you're able to deploy CloudFront instance in front of the server. With two-tiered support, AWS Shield can provide comprehensive protection against DDoS attacks small and large that is catered to your budget and needs. If you have a small product, you can get started with the standard protection and as it scaled, you can upgrade your protection to suit your needs.

Inspector

Amazon Inspector. If you or your company develops applications, there's a service that can provide your auditors and your development team a peace of mind knowing that the applications adhere to security standards set by the company and the industry as a whole. Amazon Inspector is an automated security assessment service for your applications deployed on AWS. This means that it helps you improve the security and compliance of these applications by automatically assessing them for exposure, vulnerabilities, and derivations from best practices. Once the assessment is completed, it generates detailed reports to help you check for unintended vulnerabilities. Security teams can then get reports validating that tests were performed. Inspector helps you reduce the risk of introducing security issues during deployment and development by proactively identifying potential issues that do not align with best practices and policies. You can define your own standards and best practices, and make sure that they're being followed. Or you can choose to utilize AWS's constantly updated standards. As the name suggests, Amazon Inspector inspects your applications to find security issues and bring them to your attention.

Trusted Advisor

AWS Trusted Advisor. Have you ever looked at your finances and thought, "Well dang. "I wish someone would tell me what to do with my money, "so I could have a comfortable early retirement?" I've tried some investments, I've tried putting money into a 401k, but I have no idea if I'm doing any of this right. Oftentimes, the people you call up about your money health are called financial advisors. For your AWS Cloud infrastructure, you have Trusted Advisor.

AWS Trusted Advisor is a service that guides the provisioning of your resources, so that you are following AWS best practices. Upon scanning your AWS infrastructure, Trusted Advisor advises you on how your infrastructure is or is not following AWS best practices based on five categories. The categories are cost optimization, performance, security, fault tolerance, and service limits. AWS then provides action recommendations to bring your infrastructure closer to best practice standards. All AWS customers have access to seven core trusted advisor checks for free. These checks are S3 bucket permissions, security groups, specific ports unrestricted, IAM use, MFA on root account, EBS public snapshots, RDS public snapshots, and service limits. For those with enterprise or business support plans, there are extended set of checks and recommendations available. On top of more types of checks, those with full trusted advisor access also get notifications through weekly updates, an ability to set up automated actions in response to alerts with Amazon CloudWatch. They also have programmatic access to the scan results via AWS Support API. AWS Trusted Advisor is a valuable ally in making sure that deployment of your AWS Cloud resources are aligned with best practices, as well as providing you customized recommendations based on proactive monitoring of your infrastructure.

GuardDuty

Amazon GuardDuty. In a perfect world, you wouldn't need to sleep so that you can be up at all times of the day and night monitoring your cloud infrastructure for threats and malicious activities. Unfortunately, we all need sleep, and some departments don't have the budget to have people staring at dashboards 24/7. Thankfully, you don't have to, because AWS Cloud has a service that stays up all day and night for you. It's called Amazon GuardDuty, and it's a threat detection service that monitors for malicious activity and unauthorized behavior to protect your AWS Cloud instance 24/7. It analyzes billions of events across multiple AWS data sources which can then send actionable alerts via AWS CloudWatch events. GuardDuty uses machine learning, anomaly detection, and integrative threat intelligence to identify and prioritize potential threats that may impact your AWS Cloud infrastructure. Best of all, you can deploy GuardDuty within a few clicks as there is no additional software or infrastructure to manage to take advantage of that protection. Amazon GuardDuty continuously monitors your AWS Cloud infrastructure, intelligently detects threats using machine learning and helps you take action immediately if a threat is found so that you and your team can have a good night's sleep knowing your infrastructure is being monitored at all times.

Study break: Reviewing security services

- Welcome to the Study Break. Let's review the security related services that may be relevant to the AWS Certified Cloud Practitioner exam. The security services we will be reviewing in this video are; AWS Identity and Access Management or IAM, AWS Web Application Firewall or WAF, AWS Shield, Amazon Inspector, AWS Trusted Advisor, and Amazon GuardDuty.

First, AWS Identity and Access Management, more commonly referred simply as IAM, is a free service that enables you to securely manage access to services and resources in the AWS cloud. The permission sets are extremely granular, helping you allow or deny access by users or other services to various resources. You can set access by using IAM to manage users utilizing granular permission sets. You could also create and manage IAM roles which has specific permission sets. You can allow entities to assume a role to do specific actions in your AWS cloud instance. In this way, you don't have to manually set up every entities permission sets, which could result in human errors and inconsistencies.

Finally, you can enable identity Federation, which will allow existing identities in your enterprise accounts. Many organizations allow identity Federation for their Microsoft active directory. Allowing employees to access their AWS cloud instance without having to create a new IAM user for every single employee. IAM allows you to have enhanced security, granular control over permission sets, ability to provide temporary credentials, flexible security credential management, inability to utilize identity Federation.

Second, the AWS Web Application Firewall of WAF, is as it sounds. A firewall service for web applications running on AWS cloud. It protects web apps from common web exploits as well as potential compromises that could force your apps to consume excessive AWS resources, which could be detrimental to your finances. It improves web traffic visibility, provides cost-effective web application protection and delivers increased security against web attacks. It's an affordable protection for your web applications that can be deployed within minutes.

Another security service is AWS Shield, which can protect your web applications from a distributed denial of service or DDoS attack. It provides detection and automatic mitigation of DDoS attacks to applications, helping you minimize the negative consequences and application downtime. There are two tiers available for customers. The standard tier is automatic, free and protects web apps against majority of common DDoS attacks. The shield Advanced tier provides 24/7 access to AWS DDoS response team and detects and mitigates sophisticated DDoS attacks with near real time visibility into events. And even provides financial protection against DDoS-related spikes in AWS resource usages. You can receive comprehensive DDoS protection catered around your budget and needs with AWS shield.

Next, the Amazon Inspector is an automated security assessment service for your AWS applications, which helps you improve security and compliance. It inspects your applications automatically assessing them for exposure, vulnerabilities, and derivations from best practices. After an assessment is completed, it generates detailed reports to help you check for vulnerabilities. Utilizing Amazon Inspecter helps to reduce the risk of introducing security issues by proactively identifying potential security vulnerabilities that do not align with best practices and policies. You can define your own standards to check against and create reports that validate that specific tests were performed.

You can continue enforcing best practices within your AWS cloud infrastructure. What the help of AWS has constantly updated standards made available through inspector. Another crucial security service is the AWS Trusted Advisor, AWS trusted advisor guides provisioning of resources to AWS cloud. So you're following AWS best practices. As the name suggests, it advises you on how your infrastructure is or is not following AWS best practices based on five categories; Optimization, Performance, Security, Fault tolerance and Service limits. It then offers recommendations to bring your infrastructure closer to standards. The Seven core Trusted Advisor checks are free and those with Business Support plans and above have access to Full Trusted Advisor checks. AWS trusted advisor provides customized recommendations based on proactive monitoring to make sure your AWS cloud deployments are aligned with best practices.

Lastly, Amazon GuardDuty is a threat detection service that monitors for malicious activity and unauthorized behavior 24/7. Utilizing machine learning, anomaly detection and integrated threat intelligence. GuardDuty identifies and prioritizes potential threats that may impact your AWS infrastructure. It's deployable with just a few clicks and helps you take action immediately against the threat. It works as a 24/7 monitoring solution to help your human infrastructure team get a good night's rest.

So, this was a quick study break review of the security related services AWS offers to help protect your cloud infrastructure. If you are unsure about your understanding of any, feel free to go back to the individual videos to get more in depth summaries of each, before moving on.

Study break: Exam tips and resources

Feeling ready? Let's go over some exam tips to help you prep for success. In this video, we will be going over contents that will be relevant to the security domain of the AWS Certified Cloud Practitioner exam. We will go over the shared responsibility model, principle of least privilege, security pillar of a Well-Architected Framework, and the security services. Let's get started.

The security domain is concerned with how to make sure your IT infrastructure hosted on AWS Cloud, is safe from both internal and external security threats. There are a few concepts related to this, such as the shared responsibility model, the security pillar of the Well-Architected Framework, and the principle of least privilege. In the shared responsibility model, the most important concept to understand is that you as the customer, and AWS as the cloud computing service provider, share the responsibility of keeping AWS Cloud secure. Customers are responsible for security in the cloud and AWS is responsible for security of the cloud. How do you make sure you're keeping the information in the cloud secure? You should only provide the least amount of access possible for any entity to make sure that no one has access to resources they are not entitled to. This concept is called the principle of least privilege.

Finally, the security pillar of the Well-Architected Framework states that the security of the cloud is composed of identity and access management, detective controls, infrastructure protection, data protection, and incident response. We also went over a few core security related AWS services such as AWS Identity and Access Management, or IAM, AWS Web Application Firewall, or WAF, AWS Shield, Amazon Inspector, AWS Trusted Advisor, and Amazon GuardDuty. You should be familiar with what each of these services do and how they contribute to helping keep AWS Cloud secure. If you are unsure about any of the concepts or services we just discussed, feel free to check back on the videos, or even the study break videos to review. A few extra minutes making sure you have the services or concepts straightened out could mean a few extra points on the certification exam.

AWS

Introduction to AWS for Non-Engineers

Module 3 Core Services

Technology domain

There are four domains in the AWS certified Cloud Practitioner Exam. They are: Cloud Concepts, Security, Technology and Billing and Pricing. The four courses in the Introduction to AWS for Non-Engineers series, follow these four domains. This course that you're watching now covers the largest domain in the certification, the Technology Domain.
There are four parts of the Domain which are:
1. Define methods of deploying and operating in the AWS cloud,
2. Define the AWS global infrastructure,
3. Identify the core AWS services and
4. Identify resources for technology support.

With almost 200 services available on 24 categories as of winter 2020 and ever expanding, the technology section of the domain might seem extremely daunting. Heck, they even have a satellite as a service. However there's good news, most of the services will not be on the AWS Certified Cloud Practitioner Exam. There's a certain subset of services that are often considered core services in AWS. And while dozens of services are announced every year, these core services don't change much. Generally, you do not have to worry about having to acquaint yourself with the newest product launch announcement from AWS re:Invent conference in Las Vegas every December. Study the services we talk about in this course and at least be familiar with the names of the other services, and you should be good to go. You can find all the products AWS offers at aws.amazon.com.products.

Aside from understanding the fundamentals of the core AWS services, we will also review concepts that make up the AWS global infrastructure, like Availability Zones, Regions and Edge locations. We will also briefly go over different ways you can deploy and operate in the AWS Cloud, like utilizing the AWS Management Console, through AWS Command Line Interface or CLI, and Infrastructure as Code. All of these words may seem confusing at first but don't worry, we'll get them squared away. We'll provide study breaks and exam study tips so you can begin preparing for the AWS (mumbles) Cloud Practitioner Exam's Technology domain.

EC2

Your manager comes by and taps you on the shoulder. "Hey, we need to get a new server for an app the dev team is creating. Can you get it set up for them?" "Sure!" You say, enthusiastically. After all, you love researching and buying new computers. How hard could this be? You soon realize that there are a lot of moving parts in purchasing a server. You need to know how much storage it needs, how much memory the dev team requires, the operating system, the bells and whistles. Even once you manage to find a server that fits all the requirements posed by the dev team, you then have to get it through all the bureaucratic ladders up and down the command chain in your team, and the engineering team. And once you thought you were finally ready to get it purchased, you find out you have to push it through finance, too. And the finance department is not happy with the huge, unexpected cost of buying an expensive server. You spend a little while convincing the finance department to approve the cost, and you order it. By this point it's already been a month since your manager initially asked you to purchase this and the dev team is getting impatient. Now, you wait for it to be delivered in a week. But wait, you aren't done just because it's delivered. Now you have to set it up and make it ready for dev team to use. All in all, it would take you months to get that server up and running that the dev team wanted. And by the time you deliver it to them, they have one thing to say. "Oh, yeah, we've been going further into the development process, and turns out we need double the amount of storage that we initially requested." Back to the first step.

There's an easier way to do this instead of spending months of your time drowning in bureaucratic headaches. You can log into AWS and spin up a virtual server, called an instance, in seconds with the exact specifications that your dev team requested. And if their needs change, you can easily adjust the existing server or just spin up a new one. This service is called Amazon Elastic Compute Cloud, or EC2, and it's one of the most widely used services in the AWS. Amazon EC2 allows you to launch applications and servers when you need them without upfront financial commitments. It's integrated with many other AWS services, is reliable and secure, and allows you or your company to quickly and inexpensively spin up instances of virtual servers for all of your different needs.

Elastic Beanstalk

Do you find yourself doing more administration of infrastructure than coding, even though you are a programmer? Are you spending more time managing your platform, provisioning, and load balancing than developing? Do you wish you could worry about your code instead of whether or not your application stack can handle your app? Or wish you could provide this environment for your team's developers? AWS Elastic Beanstalk could be your solution, Elastic Beanstalk is an easy to use AWS service to help you deploy and scale web applications by simply uploading your code.

Elastic Beanstalk handles the deployment process, including the capacity provisioning, load balancing, auto scaling and application health monitoring. You can upload services developed using JAVA, .NET, PHP, Node.js, Python, Ruby, Go and Docker and you retain full control over the underlying resources at all times, it's free to use and you only pay for the AWS resources needed to store and run the web applications you've deployed. You never have to worry about outgrowing the resources Elastic Beanstalk provisioned for you because it automatically scales your applications up and down based on its specific needs. You also have complete freedom to select the AWS resources such as the EC2 Instance type that you want to use for your application. If you decide that you want to take over the manual management of the infrastructure you can do so at any time. Elastic Beanstalk provisions and operates the infrastructure for you so that you can focus on coding. If you're constantly frustrated by the amount of time you spend managing configuring servers, databases, firewalls and networks, perhaps its time you give AWS Elastic Beanstalk a try.

Elastic Load Balancing

Have you even experienced a situation where suddenly everyone around you at work seems to want something from you? And now you probably got completely overwhelmed and maybe found yourself unable to do any of the tasks requested much less all of them. When a server is overloaded with requests, it reacts similarly. It becomes unable to send out the responses to the paralyzing amount of traffic. It becomes completely overwhelmed and you might notice that the website is loading very slowly or goes down completely.

Let's think about fashion company A. Fashion company A decided to have a sale to celebrate their first anniversary. All shirts are 50% off. Their marketing department did an amazing job and everyone is talking about it on social media. As soon as the sale starts, boom, their site goes down. Uh oh, what happened? Chances are the infrastructure team wasn't prepared for the crushing amount of traffic that suddenly inundated the company website server which quickly became overloaded. Now the whole company is running around trying to get the website back up and communicate with angry customers. How could fashion company A have prepared better for the sale so that their website didn't go down from too many excited customers accessing it at the same time? They could have utilized elastic load balancing to automatically distribute incoming traffic across multiple servers. This means that the infrastructure department could have set up multiple replicated servers which means multiple servers with the same content on them and place the elastic load balancer in front of them to distribute the traffic to multiple servers. That way each server is only taking care of a fraction of the overall traffic. You can think of elastic load balancers like air traffic controllers telling incoming airplanes to go to different runways. They make sure that any one runway doesn't get overwhelmed with airplanes to cause delays by equally utilizing all available runways for landing. Elastic load balancers help your applications achieve fault tolerance by ensuring scalability, performance, and security. They can also monitor the health of your servers. If a server goes down, the load balancers can send the traffic to the remaining healthy servers. Elastic load balancers are highly available, secure, flexible, and monitorable allowing you to glean robust information about your traffic as well as providing you the confidence that your applications are up at all times.

Lambda

You've created a mobile application that allows people to create profiles of dogs and friend other dogs all over the internet. The first thing a new user does is upload their dog's photo, because what's a dog social media without dog photos, right? You want the dog photo's thumbnail to become available immediately after upload, so that the user can pick and choose which photos to feature on their profile. In order to make this happen, you need to provision and manage servers to run your code and make your app function. As more dog lovers sign up, you find yourself having to devote more and more time to maintaining and scaling your server infrastructure, instead of writing code to make your app better and adding new features. How can you focus on writing code for your dog social media instead of spending so much time on the infrastructure upkeep? You can check out AWS Lambda.

AWS Lambda runs code, called a Lambda function, in response to an event. An event could be anything from a user uploading an image into an S3 bucket to a user tapping a button on your mobile app to buy an item. All you have to do is upload your code, and AWS Lambda automatically runs your code and scales your application for you. You only pay for the time your code spends running, and each event trigger, which helps you keep your costs low. There are no servers to provision or manage, and nothing to scale, because Lambda takes care of all of that for you so you can focus on writing code. You can have a Lambda function up and running with just a few clicks, saving you hours and hours of back-end provisioning work. More dogs, less admin work.

Lightsail

Love the idea of having your website, database, or application running on AWS, but just don't have the energy or technical know-how to get it done? Need a quick development or test environment spun up? AWS has a service that'll do just the trick, Amazon Lightsail. Amazon Lightsail is perfect for simpler workloads, quick deployments, and getting started on AWS. It's a snap to get going, while still being designed to scale with you as you grow. You can use it to deploy simple web applications, create websites, run your business's software, or spin up developer or test environments, while maintaining cost-effective monthly fees. There are many preconfigured and ready-to-use operating systems, web apps, and development stacks.

Some of the more popular resources available are WordPress, Windows OS, Ubuntu, and Node.js. They are one-click-to-launch services, so getting started is a breeze. You can quickly deploy projects ranging from creating your first WordPress blog to running a database with just a few clicks. If you're considering using AWS to quickly spin up projects or resources, but don't have the time or engineering know-how to deploy full-on services, consider trying out Amazon Lightsail.

Deploying and operating in AWS

In this video, we're going to take a break from all the services, and discuss an important topic when it comes to creating and maintaining your IT resources on the AWS Cloud, deploying on to AWS Cloud. AWS users can manage and deploy resources since the AWS cloud in three ways, by utilizing AWS Management Console, Command-Line Interface, or CLI, and Software Development Kits, or SDKs. They all reference the AWS API's to help you deploy, and manage your AWS Cloud infrastructure.

The AWS Management Console is a graphical interface that supports the majority of AWS services. You can think of this as a web portal that you log into, as you would your social media account to see everything that's offered on that website. Through it you can look at your billing statements, launch new services, and find out how your apps are doing, all while using a interface that feels like you're browsing through another website. If you don't have much familiarity with utilizing command lines, or SDK's, the AWS Management Console is extremely user-friendly and easy to navigate. The AWS Command-Line Interface, or CLI, allows you to access services via the command line. The command line is a way to access and change resources with text-based command entry. You can tell an easy two service to shut down, or add new file to an S3 bucket, all by typing in some command into the AWS CLI. It is programming language agnostic, and allows you to create scripts to run on AWS.

The AWS Software Development Kits, or SDKs, let's you incorporate connectivity and functionality of a wide range of AWS cloud services into your code, helping you to deploy AWS services and resources using a variety of popular programming languages like NODEJS, C++, Java, Ruby, and PHP. SDK allows you to use AWS Cloud resources in existing applications. You can use just one, two, or all three ways of deployment to AWS Cloud. AWS CLI's and SDK's allow you to create tools that are specific to your organization, and help create an environment that utilize infrastructure as code. With infrastructure as code, you can write code that describes the configurations for specific AWS Cloud services, and they can be deployed for you by AWS. It helps to speed up the deployment process, and removes the risk of human error when spinning up new resources. Some AWS Cloud services that utilize infrastructure as code are Elastic Beanstalk, AWS Lambda, and AWS CloudFormation.

AWS Global Infrastructure

Amazon web services is global. Many companies and users around the world rely on AWS cloud to help their businesses succeed and grow. AWS has data centers around the world called Availability Zones. Each availability zone is independent of each other in network and power source. There are currently almost six dozen availability zones, or Azs, around the world. A region is made up of two or more availability zones, and there are currently two dozen AWS regions around the world. Some regions have more AWS cloud services than others. When a brand new service is introduced, it's generally first introduced in a few specific regions as opposed to the whole world. Some of the regions that receive new services earlier are US East North Virginia, US West North California, some Asia Pacific regions like Singapore, Sydney, and Tokyo, and some areas in the European Union like Frankfurt and Ireland. This fact may influence which region you decide to use to host your infrastructure.

Generally, you would choose a region closest to your physical location to host your AWS cloud infrastructure, because you can reduce network latency for your end users. For example if your company is based out of Washington D.C., you might pick the region US-East-1, which is based out of North Virginia in the United States. Some regions cost more than others. And service level agreements, or SLAs, also vary by region. You may also have compliance requirements to meet, which may require you to host your resources in specific regions or multiple regions. Hosting your resources in multiple availability zones, or even regions, help create what is known as high availability. The ability to provide uninterrupted performance even during natural disasters is called resiliency. Having multiple copies of your data in different data centers is called having redundancy. By architecting you AWS cloud infrastructure across multiple regions, you can prepare for events like power failures, natural disasters, and other potential operational mishaps to make sure that your web applications and resources don't experience extended down times when disaster strikes. Every data center, availability zone, and region is interconnected with highly available, low latency, private global network. This means that data transfers between data centers, availability zones, and regions are super fast, allowing customers to take advantage of the resources without lag.

Study break: Reviewing compute services

In this chapter, we went over five of the major compute services in AWS. Amazon Elastic Compute Cloud, or EC2, Amazon Elastic Beanstalk, Elastic Load Balancing, AWS Lambda and Amazon Lightsail. Let's quickly review all of them to make sure we've got the fundamental concepts down before moving on.

Amazon Elastic Compute Cloud, more commonly referred to as Amazon EC2, is a virtual server hosted on AWS cloud. You can instantly launch applications and servers wherever you want with an extremely versatile range of capabilities. It is one of the most widely used services in AWS and you can spin up an instance with no upfront financial commitments.

AWS Elastic Beanstalk helps you deploy and scale web applications by simply uploading your code. It handles the deployment process like capacity provisioning, load balancing, auto-scaling, and application health monitoring so you and your team can focus on coding. You can upload code in many of the popular programming languages like PHP, Node.js, and python. And you can retain full control over the underlying resources at all times. Better yet, it's free. You only pay for the AWS resources utilized as a result of deploying the code.

Elastic Load Balancing helps your application achieve fault tolerance by ensuring scalability, performance, and security. They can monitor the health of your servers, and if one goes down, it can reroute incoming web traffic to healthy servers. Elastic Load Balancers are highly available, secure, flexible, and monitorable which means you can feel confident that your applications are up at all times and even get insightful information about web traffic going to your applications.

AWS Lambda runs code called a lambda function in response to an event. An event could be someone uploading an image into your application, or someone visiting a specific web page. When an event happens, your code runs. Instead of having to spin up and maintain servers to take care of set events, lambda allows you to deploy the lambda function and only charges you for the time your code spends running in each event trigger. No servers to provision, manage, or scale. Lambda functions can be up and running with just a few clicks, saving you tons of time.

You can think about *Amazon Lightsail* as a EC2 on bumper bowling mode. AWS provides many preconfigured and ready to use operating systems, web applications, and development stacks to help you get your applications or websites up and running with minimal configurations. When you want to focus on launching and creating instead of configuring and managing your virtual servers, Amazon Lightsail might be the better option to you over Amazon EC2. It's a snap to spin up a virtual server, but the servers are still designed to scale with you with cost effective monthly fees. Some popular resources available to you are Word Press, Windows OS, Ubuntu, and Node.js. With a one click to launch service like Amazon Lightsail, there's no real excuse to not get your project up and running on the cloud.

Of the services we went over, AWS Elastic Beanstalk and AWS Lambda are what we consider infrastructure as code because they allow us to deploy resources to the cloud using code.

There are three ways to deploy to the AWS cloud.

1. The AWS Management Console,
2. AWS Command-Line Interface, or CLI, and
3. AWS Software Development Kits, or SDKs.

Finally, AWS has a global infrastructure.
They have independent data centers called Availability Zones.
Two or more Availability Zones make up a Region. Your AWS cloud infrastructure is generally hosted in a Region closest to your organization's physical location. By creating redundancy by replicating your resources in multiple Availability Zones or Regions, you can create an infrastructure that is resilient to natural disasters and highly available.

S3

I have my files, folders and videos backed up to the cloud everywhere, Dropbox, Box, Google Drive, you name it, I probably have some files stored there. Most of the time it's because I've run out of storage in one service, so I move to another, but this makes it so difficult to keep tabs on all of my files, and I could lose important documents if I can't get into my account or remember which account I stored them in. Many of us have had hazard file storage situations when it comes to backing up to the cloud. The storage service might go down, we might be wasting money on paying for storage space through a subscription service that we may not end up using. Also, when we upload the files to the cloud, it's difficult to hotlink to it, meaning you can't use an image or upload it onto a service like Dropbox or Google Drive and embed it into your webpage.

AWS has a solution to all of your static file storage called Amazon Simple Storage Service or Amazon S3. Amazon S3 is an object storage service, which means that you're storing each file as an entity called an object. It offers industry-leading data availability, security, performance and scalability. Scalability refers to the way you can scale your usage up or down with extreme flexibility and be charged only for what you use. It's designed for 99.999999999% or 11 nine's of durability, which means that there is almost no chance of the data become corrupted. You can upload files of all sizes to serve a wide variety of needs such as websites, mobile apps, backup and archiving, enterprise applications, IoT devices, and big data analytics, as three boasts easy-to-use management features to fine-tune access controls for your organization's specific compliance requirements.

There are many storage classes available with S3, which support different data access levels at corresponding pricing. You can even setup S3 lifecycle policies, which would automatically transfer files from one storage class to a cheaper one after a certain number of days. These options range from using S3 Standard class to store your frequently accessed files to using S3 Glacier Deep Archive to store backup data that's rarely accessed for very cheap rates. Whatever your needs for object-based files storage may be, ranging from using photographs for your website or spending as little money as possible to backup your organization's files, Amazon S3 would have an option that works for your budget and needs.

Elastic Block Store

- You've span up a virtual server on AWS, using amazon EC2 to run a database. But you notice that you are running out of space on your virtual machine. What should you do so that you can continue making your databases larger without impacting the virtual machine's performance? AWS has a solution for you called Amazon Elastic Block Store which allows you to add extra block stores to your EC2 instance and you don't even have to reboot your server.

Amazon Elastic Block Store or Amazon EBS behaves like raw , unformatted blocks devices, which can be mounted or attached to your EC2 instance to expand your server storage. You can add multiple volumes to the same instance, and you can use these volumes as file systems, or hard drives. You can dynamically change the configurations of a volume attached to an instance, which means you can change the settings and sizes with just a few clicks on the management consule. These volumes are automatically replicated within their availability zone, making them available and durable.

Many organizations use EBS to host their huge databases. There are different EBS stores types available to fit your needs and budgets as well as the option to encrypt the volumes for compliance. EBS provides persistence block storage volumes which means that they don't disappear when EC2s are rebooted. They also exist independently of the virtual servers they are mounted on to, and can therefore be moved around on to other EC2 instances. You can think of EBS volumes as external hard drives for your virtual servers, taking advantage of scalable, durable, and reliable storage options using Amazon EBS will make scaling your IT operations a breeze.

Snowball

Okay, you've made your case and management has approved the migration of the company's backups from onsite servers to AWS Cloud. Here we go! 10 years worth of data. Time to upload. Wait, how long will uploading 10 years worth of data even take through the Internet connection at your company? Sometimes the amount of data you want to upload to AWS Cloud is small enough that you can do it through your high-speed Internet connection. Other times, you realize that the amount it takes for you to upload gigabytes or terabytes of data would potentially take years on the Internet circuit you have. But, of course, AWS has already come up with a solution to make your life easier. AWS Snowball.

AWS Snowball is one of the very few hardware solutions from AWS, and it is a data migration tool. This means that AWS will physically ship you a Snowball to move your data onto, and then mail it back so you can migrate huge amounts of data. The amount of data you can move into AWS Cloud ranges from 50 terabytes with regular Snowball up to 100 petabytes with a Snowmobile, a 45-foot-long ruggedized shipping container pulled by a semi-trailer truck. In case you had any doubts, this is the Snowmobile, pulled by semi-trailer truck to haul your petabytes of data back to AWS Data Centers. The usage fee of a Snowball device is free for 10 days of onsite usage, with small extra onsite usage fees for every day you keep it beyond that. There is a service fee per job which ranges from $200 for a 50 terabyte snowball to $320 for an 80 terabyte snowball used in Asia-Pacific. The data transferred into AWS is stored in S3 and there is no transfer fee. However, you will be paying for S3 storage space as a storage fee, to host your transferred data.

The Snowmobile has larger costs associate with use as you will be chartering a huge truck with extremely large storage capacities. You can follow up with an AWS sales representative for an estimate for your particular need. Requesting and using AWS Snowball is simple. You make a request to have the hardware mailed to you using the AWS Management Console. When it arrives you attach the device to your local network, run the Snowball client on your machine, and select folders and files you want to encrypt and transfer onto the AWS Cloud. Once the transfers are completed you can mail it back to AWS and once received your files will be uploaded onto S3. 100 terabytes of data will take more than 100 days to transfer over a dedicated 100 megabits per second connection. For large amounts of data, the transfer up to the cloud could take months, even with a high speed Internet connection. With AWS Snowball the same transfer can be accomplished in less than a week using two Snowball devices, with a few days tacked on for shipping. If you're considering moving a lot of data onto AWS Cloud, take a peek at AWS Snowball to help you out.

Storage Gateway

So, this idea of using cloud computing to keep your costs down sounds like a great idea but your organization uses the data a lot so you also don't want to sacrifice latency or the length of time it takes for your resources to be accessible. Going up to the cloud to download a 500 megabyte file every time you want to make an edit sounds like a horrible idea when your whole company is sharing a single circuit. What should you do to maintain very low latency but still take advantage of the costs and time-saving benefits of cloud computing? AWS Storage Gateway may be the best of both worlds solution you're looking for. It connects your on-premises storage with AWS Cloud Storage, providing a hybrid storage solution for your IT infrastructure.

The service seamlessly integrates on-premises enterprise applications and corporate workflows with AWS's Cloud Storage Services through the use of a virtual machine installed onto an on-premises data centers host server. Basically, it creates a gate that connects your on-site users and devices to the resources stored in AWS Cloud with minimal latency.

AWS offers three types of storage solutions to fit your needs.

1. File based,
2. volume based, and
3. tape based.

Files backed up using the *File Gateway* are stored as objects in S3. There is a one-to-one representation of each file backed up to the cloud in S3 and the gateway asynchronously updates the objects in S3 as local files are updated. Local cache is maintained to provide low-latency access to recently access resources.

The *Volume Gateway* on the other hand, uploads files in blocks as opposed to single files. You can think of the Volume Gateway as backing resources up as a virtual hard disk instead of individual files. These blocks can be asynchronously backed up as point-in-time snapshots and stored as Elastic Block Store or EBS snapshots. There are two types of Volume Gateways available:

1. stored volume, and
2. cached volume.

The major difference is where the complete copy of your data is stored. Stored volume keeps the complete copy on-premises while sending snapshots or incremental backups to AWS. Cached volume keeps only the most recently accessed data on-premises and keeps the complete copy on the cloud.

The last type of storage gateway is the *Tape Gateway* which utilizes virtual tapes. You can use your existing tape based backup infrastructure to backup data onto virtual tapes on S3. You can think about Tape Gateway as taking backups on physical tapes except instead of physical tapes they are digital tape cartilages stored on S3. Data is stored locally then asynchronously uploaded to S3. The data can then be archived using Amazon Glacier which is like sending your physical tape backups to an off-site tape holding facility like Iron Mountain. You pay for storage and data retrieval. The quicker you can access the backed up data the more expensive the solution becomes. For example, data stored via Tape Gateway is much cheaper saved to S3 Glacier Deep Archive than S3 Glacier because the data retrieval takes a longer period of time.

Depending on where you want to store the complete copy of your data and how you would like to backup your data there are multiple options available for utilizing AWS Cloud as a backup and storage resource for your frequently accessed data through AWS Storage Gateway.

Study break: Reviewing storage services

In this chapter, we run over four of the major storage services in AWS, Amazon Simple Storage Service or S3, Amazon Elastic Block Store or EBS, AWS Snowball, and AWS Storage Gateway.
Let's quickly review all of them to make sure we've got the fundamental concepts down before moving on.

Amazon Simple Storage, more commonly known as Amazon S3, is an object storage service, which you can conceptualize like storing each file as an individual object, like you would in your My Documents folder. It's designed for scalability, data availability, security, and performance, and is used in industries of all sizes. There are many storage classes available to fit every organizations budget and needs. You can even set up S3 lifecycle policies which will automatically transfer files from one storage class to a cheaper one after a certain number of days. You can use S3 for a variety of needs whether it's for hosting images your users upload to your web app as your as inexpensive backup solutions.

In contrast, *Amazon Elastic Block Store or EBS* is a block storage service while S3 stores files individually as objects, Amazon EBS stores them as blocks. They're not a 100% accurate representation, you can think of it like making a zip file of your Christmas holiday photos and sharing that zip file with a friend. They can't just download a portion of the zip file, they have to download the whole thing to unzip it and see your individual photos. Amazon EBS behaves like raw, unformatted block devices, which can be attached to your EC2 instances to expand your server storage. Think about an external hard drive that helps you up your laptop storage capacity. Amazon EBS is scalable, durable, and reliable storage option to make sure you always have enough storage available for all of your applications and servers.

AWS Snowball is one of the very few hardware services AWS offers. It's a data migration tool that can also function as a storage service. When you begin using AWS Snowball, AWS will actually ship you a physical Snowball to move your data onto. Once you finish moving the data onto Snowball, you mail it back and AWS will migrate the data onto Amazon S3 for you. The amount of data you can transfer to AWS cloud at one time using this service ranges from 50 gigabytes with a regular Snowball to up to 100 petabytes with a Snowmobile, a 45 foot long shipping container pulled by a semi-trailer truck. Why bother with a physical device? Because transferring such large amounts of data over the internet will take a lot of time. By moving the data to a physical device and then shipping it to AWS to upload to S3 on their end, you can save a lot of time, bandwidth and even money.

AWS Storage Gateway is a hybrid storage solution for your IT infrastructure providing both low latency for file access and benefit of cost and time saving with cloud computing. It is a gate that connects your onsite users and devices to resources stored in the AWS cloud with minimal latency, and offers three types of storage solutions to fit your needs, file gateway, tape gateway, and volume gateway, all addressing different kinds of needs. In the most fundamental sense, the difference in the three gateways is where you want to keep the complete copy of your data, onsite or on the cloud. There was a lot to unpack with the storage services. They all store things for you in the cloud, but they all have different uses in different ways of getting your data there. If any of them seem like they need a little bit more reviewing, feel free to go back and review the videos.

DynamoDB

Your team has created a mobile application and plans on releasing it next week. And a prominent TEC Publication has picked up the story. The publicity has done wonders and the application has 100 times the number of anticipated downloads, that's fabulous. But what's not fabulous is that your back and infrastructure is not prepared for the amount of data that's going to be sent back and forth, and your database solution can't scale up quickly enough to meet the new demands, while maintaining the efficiency your audience is looking for. Thankfully, AWS has a low maintenance solution for your database that scales up and down with your needs.

Do you need a secure scalable, fast and flexible NoSQL database?
Amazon DynamoDB is a one stop solution for enterprise ready database needs that can help you build applications with virtually unlimited throughput and storage.

This means that no matter how small or large your project is, DynamoDB will adapt with your needs so that your application speed and stability remains high. DynamoDB is serverless which means that you don't have to provision, patch or manage any servers. AWS automatically scales your tables up and down to adjust for capacity and to maintain performance as well as maintaining stability with redundancies and fault tolerance. You can focus on getting rid of that pesky bug from your new Superstar mobile application, instead of provisioning and maintaining the databases. You have the option of paying only for what you use, or specifying a workload amount and paying for the provision resource depending on your organization's preference.

DynamoDB is being utilized in many well known companies like Nikee, Snapchat Lyft and Netflix. your organization's application may benefit from utilizing AWS's scalable NoSQL database service.

RDS

You just learnt about Amazon solution to a non-relational database: Amazon DynamoDB. But what if your application requires a relational database, and an efficient yet affordable one at that. As you'd imagine, AWS has a solution for that as well, and it's called Amazon Relational Database Service, or Amazon RDS. Don't you love it when the name of the service tells you exactly what it's meant for?

With Amazon RDS, you can set up, operate, and scale a relational database with just a few clicks. It's cost-efficient, and you pay only for the resources you actually consume. You also have the option of on-demand pricing or even lower hourly rates with Reserved Instance pricing. Because you don't need to be responsible for provisioning, monitoring or maintaining the database, you can focus on building your application secure in the knowledge that AWS make sure your databases are highly available, durable, scalable, fast and secure.

Best of all, while AWS does provide its own database engine called Amazon Aurora as an option, you can utilize very familiar database engines like PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server. So you can dive right into creating a database without the learning curve. If you already have existing databases, you can use AWS Database Migration Service to migrate or replicate them into Amazon RDS as well. In the next video, let's check out AWS' own relational database engine, Amazon Aurora, which is fully managed by Amazon RDS.

Aurora

Amazon Aurora is one of the six relational database engines that you can use in Amazon Relational Database Service, or RDS. It's fully managed by Amazon RDS, which relieves engineers of having to engage in time consuming administrative work, like provisioning, database setup, and maintenance. Instead, it allows the engineers to focus on coding.

You can monitor performance using various AWS monitoring and alerting services so you can quickly detect performance issues. It's MySQL and PostgreSQL compatible, but up to five times faster than standard MySQL and three times faster than standard PostgreSQL databases. You can get the same security, availability, and reliability of commercial databases for just 1/10 of the cost. You can have databases scaling all the way up to 64 terabytes per instance, hosted on distributed, fault-tolerant, self-healing storage systems with low latency. You can migrate existing MySQL or PostgreSQL databases onto AWS using AWS Database Migration Service. Inexpensive, highly scalable, available, durable, and secure, Amazon Aurora is a great relational database service to look into.

Redshift

Got lots of data? Petabytes of data? Looking for a data warehouse to efficiently and affordably analyze it all? Well, of course AWS has a service to help you manage all that data, and it's called Amazon Redshift.

Data warehouses are large stores of data accumulated from a wide range of sources within a company, which is then used to guide management decisions after analysis. Amazon Redshift is a cloud-based, fully managed, petabyte-scale data warehouse service. Always striving to improve, Redshift in 2019 is 10 times faster than it was just two years ago. You only pay for what you use, and it starts as low as 25 cents per hour, all the way up to petabytes for under $1,000 per terabyte per year. Redshift is fully integrated with your data lakes, which are repositories of data stored in their raw formats. You can deploy a new data warehouse in minutes, and it's easily scalable with the click of a button. Security is built in too, so you can secure your whole database with the click of a mouse. Data encryption is also simple, and Redshift's encryption is compliant with many common requirements like HIPAA. If you're looking for a data warehouse solution that's cheap, secure, and easy to deploy and manage, Amazon Redshift might be a good place to start.

Study break: Reviewing database services

In this chapter, we went over four of the major database services in AWS. Amazon DynamoDB, Amazon Relational Database Service, or RDS. Amazon Aurora, and Amazon Redshift. Let's quickly review all of them to make sure we've got the fundamental concepts down before moving on.

Amazon DynamoDB is a fast, flexible, fully managed and secure non relational, or NoSQL. Database, that can handle more than 10 trillion requests per day, and support peaks of more than 20 million requests per second. It's serverless, so you don't have to provision, patch or manage any servers, and it automatically scales up or down, to adjust for capacity. Instead of worrying about managing your database, you can just worry about scaling your application.

Amazon relational database service, or Amazon RDS is a fully managed relational database, because it's fully managed, like Amazon DynamoDb. You don't have to provision, or manage any servers. Instead of spending your time doing administrative tasks, you can devote your time to working on your products.

You have six database engines to choose from:

1. Amazon Aurora,
2. PostgreSQL,
3. MySQL,
4. MariaDB,
5. Oracle database, and
6. SQL Server.

We just mentioned *Amazon Aurora*, as one of the database engines you can use with Amazon RDS. Amazon Aurora is fully managed by Amazon RDS, and it's MySQL and PostgreSQL compatible. You can get the same security, availability and reliability of commercial databases, but for faster and cheaper. *Amazon Redshift* is a cloud based, fully managed, petabyte-scale data warehouse service, that's faster and cheaper than other data warehouse providers. Data warehouses store extremely large amount of data, collected from a wide range of sources to analyze. It is quick to set up, and easy to scale, and its encryption, is compliant with many industry regulations.

Databases, databases. You may not be very familiar with databases, which could make trying to decipher these options a little more difficult. *Amazon DynamoDB,* is a **non** relational or NoSQL database. Whereas Amazon RDS, and Amazon Aurora, are relational databases. Amazon Redshift, is a data warehouse for lots and lots of data. All are scalable, secure, and less expensive than industry alternatives. If you want to review any of the databases, feel free to take a few minutes, going over to videos.

VPC

Think about your wi-fi setup at home. That's a private network. You likely have a cable that runs into your house from the street, which connects your home's private network to your ISP, like Comcast or Verizon. That cable is connected to your modem, which is your connection to the internet, also thought of as a gateway. The modem is connected to a router, or a switch, via another cable which routes traffic between devices in the network, and also the internet. You connect your devices, like your laptop and tablet, to the router, using the wireless network. Your home wi-fi setup is a private network where you can create your own ecosystem for connecting devices and resources. And a private network in the cloud is what AWS calls Amazon Virtual Private Cloud, more commonly referred to as Amazon VPC.

Amazon VPC creates a logically isolated section in the cloud where you can provision your AWS resources. Think of it like your corner of the cloud, where you define what comes in, what goes out, and what lives inside. Amazon VPC is very flexible and secure, allowing you to control almost every aspect of your virtual network. It's completely scalable, allowing you to instantly scale your resources up or down. It also boasts advanced security features, like security groups and network access control lists to help you filter inbound and outbound traffic at the instance level and sub-net level.

When you sign up for an AWS cloud account, you automatically get a VPC provision to you, along with automatically configured sub-nets, IP ranges, route tables, and security groups to help you get started. Going back to the analogy of your home network, the virtual private cloud is your home network. The modem is the internet gateway. The router is the route table. And your network's firewall is the network access control list. Your laptops and tablets are resources like your EC2 instances that are launched inside your VPC, or private network. When you create your first AWS cloud account, you will be creating a logically isolated corner in this vast realm of the cloud, where you are free to create and scale resources for your organization.

CloudFront

Faster, faster, faster. That seems to be the trend with everything these days. Back in the days, we were happy just to have objects we ordered online reliably delivered to us. But then Amazon's two-day shipping happened and suddenly waiting a week for your shoes to arrive became almost intolerable. In some cities, Amazon even has one-hour shipping for that emergency toilet paper you just can't seem to leave the house to buy.

In the digital space, we went from having to buffer a 10-minute YouTube video to being able to watch a whole 4K movie with no apparent lag on Netflix. Gone are the days when you had to wait for images to load slice by slice or a whole webpage to load a few rows of text at a time. We want things fast, digital or physical, and we want them now.

On the Internet, content delivery networks or CDNs are working behind the scenes to deliver your content faster and faster. Amazon's global CDN service that securely delivers data, applications, and APIs is called Amazon CloudFront.

CloudFront seamlessly integrates with many AWS services to provide optimal performance and security, including AW Shield for DDOS mitigation and Amazon EC2 as origins for your applications. A CDN is a system of distributed servers around the world that delivers website and application content to end users based on a few factors. These factors are location of the user, origin of the website or application, and the location of the content delivery server. The main purpose of CDNs is to make loading websites and applications for end users faster. And Amazon CloudFront does this by using edge locations to cache files and resources for quicker retrieval.

Imagine your favorite fruit. Most of us go to the grocery store to pick up an apple or a watermelon. Not very many of us live close enough to a farm to directly buy a fruit from a farmer. Instead, there are distribution networks set up between us and the farmer, which simplified bring the fruit to a grocery store nearby. All we have to do is pick up the fruit from the local grocery store where a truck traveled for days to bring the fruits from the farm. We have the convenience of driving 15 minutes as opposed to hours or maybe even days to buy some fruit. The farm is the origin, which on AWS could be S3 bucket, EC2 instance, or Elastic load balancer amongst a few other services. The truck then takes the apples to a grocery store where they are left to be sold to consumers or, in web terms, cached.

In AWS, files and data are cached at edge locations. Once the data is downloaded to an edge location it stays there for a certain period of time at which point users near the data center can retrieve the webpages or application resources from the edge location close to their location rather than having to go all the way to the origin which could even be on a different continent. This allows for data to be retrieved faster with the best possible performance because users are not going all the way back to the origin server to download the resources but rather accessing a location close to themselves. CloudFront is scalable, allowing you to start small and scale up as traffic to your application or webpage increases. It automatically manages traffic load without any intervention from you and utilizes application acceleration and optimization. There is no minimum commitment or a fixed-term contract and you only pay for content delivered using the service. Amazon CloudFront acts as a supermarket in a busy city making cached data quickly accessible to users around the world using edge locations.

Route 53

If you've ever set up a website with yourpersonaldomain.com, you probably used a domain name registrar to purchase and set up your domain. You might have used something like GoDaddy or Namecheap, to name a few popular commercial domain name registrars. AWS, of course, has its own service where you can purchase and set up domain names, but it can do so much more. It's called Amazon Route 53, and it's a highly scalable cloud Domain Name System, or DNS.

It allows you to reliably and cost-effectively route your end users to your internet applications. It can connect user requests to infrastructure running on AWS, like an EC2 instance or an S3 bucket. It can also route users to infrastructure outside of AWS, acting as a DNS service for domains purchased at other domain name registrars.

Route 53 is designed to be integrated with other AWS services, like mapping your domain names to your EC2 instance or S3 bucket. It's simple to set up, fast, secure, and cost effective. You are charged only for what you use, without any upfront fees or minimum usage commitments. It's also designed to automatically scale to handle large query volumes. Route 53's basic functions are domain registration, domain name system, or DNS, service, health checking of web application accessibility, and auto naming for service discovery. Utilizing the more robust features of Route 53 allows you to create websites and applications with high availability by automatically rerouting traffic catered to demand and integrating with services that send alerts when down times occur.

One of the ways you could use Route 53 is to purchase yourpersonaldomain.com, then to route users who visit yourpersonaldomain.com to a static website, hosted on your S3 bucket. When your visitors type in yourpersonaldomain.com into a browser, they will be able to load your static website, because Route 53 does the legwork of routing the user to the resources you identified. You can think of Route 53 like a telephone operator. Back in the days, you would call the telephone operator so you could speak to your grandmother. When the telephone operator got your request, he or she would use a switchboard to connect you to your grandmother. Route 53 works in similar ways, except with routing internet traffic.

Study break: Network and content delivery

In this chapter, we went over three of the major network and content delivery services in AWS.

1. Amazon Virtual Private Cloud or VPC,
2. Amazon CloudFront, and
3. Amazon Route 53.

Let's quickly review all of them to make sure we've got the fundamental concepts down, before moving on.

Amazon Virtual Private Cloud or VPC is an isolated corner of AWS Cloud made just for you. You can provision your AWS resources into a virtual network that you define with complete control over your virtual networking environment. From IP address range, to configurational route tables, and network gateways. It's free, and it's automatically created for you when you create your AWS account. Inside your very own Amazon VPC, you can create and scale your AWS Cloud resources to your heart's content.

Amazon CloudFront is a content delivery network or CDN. The main purpose of CDNs is to make websites and applications load faster. Amazon CloudFront achieves this by using Edge Locations all around the world to cache files and resources for quicker retrieval. By caching, say a video, at an Edge Location in Sydney Australia, someone who lives in Australia can stream the video much quicker than if there was no content delivery networks. Because they would have to download the video all the way from the content origin, which could be anywhere in the world. Amazon CloudFront sees where you're based and routes your traffic to the closest cache location. So you can enjoy the content without having to wait. It's scalable, and you only pay for content delivered using the service.

Amazon Route 53 sounds like a highway, and in a sense, it is kind of like a highway. If highways help take you from here to there, it is a highly scalable domain name system or DNS. It allows you to route your users to your internal applications. This could be in form of your users accessing infrastructure running on AWS, like an EC2 instance. It's basic functions are: domain registration, Domain Name System or DNS, health checking of web applications' accessibility, and auto-naming for service discovery. It helps route your users to the appropriate resources you want them to access. AWS's networking content delivery services mainly occupy themselves with helping you create secure networks for your resources to live within, and to route traffic to proper services. If any of these services we just talked about needs a bit of review, feel free to go back and check out the videos again.

CloudFormation

You've built an awesome system in AWS using many different services and many different settings. Together, they work great, and you'd like to replicate the setup for a new project. Unfortunately, setting it up take days and you don't exactly remember every step you took. Worse yet, some resources have to be provisioned before others for the system to work. We love recipes when we cook because they tell us what materials to buy, and when to do what for that perfect meal. In tech, we also love recipes. Better yet, we love recipes that cook themselves and present us with the finished dish.

That's what Amazon CloudFormation does for your IT infrastructure hosted on the cloud. You create templates, like recipes for your resources to be set up a certain way. And you can run it over and over to provision and deploy fully configured infrastructure. Best yet, unlike all those cookbooks filled with your favorite recipes in the bookstores, using CloudFormation is free. You just pay for the resources you use when you run the service, like the EC2 instances or S3 bucket storage.

With CloudFormation you can provision anything ranging from a simple EC2 instance, to a multiregion multitier application quickly and efficiently, using a simple text file written in JSON or YAML. You can update or manage the templates, referred to a stacks at any point using the AWS Management Console, command line or Software Development Kit, commonly known as SDK. Basically, you can change up the recipe whenever you see fit, even making different versions for different uses.

Version control is always available so you can revert back to previous settings whenever you want. AWS CloudFormation brings to life what is known as Infrastructure as Code where you can deploy IT infrastructure based on a text file filled with code, that specifies resources and configurations you need for each service you want to deploy. With CloudFormation, you can bring orderly and predictable back into resource deployment, no longer leaving things up to human error or chance.

CloudTrail

Your AWS IT infrastructure, like any IT infrastructure, needs to be monitored and audited, to make sure that the resources remain compliant with any government, industry, or company policies. In addition to compliance, the AWS CloudTrail service, helps to track user activity and API usage. Which allows for operational and risk auditing of your AWS infrastructure.

With CloudTrail, you can log and monitor account activities, provide event history of account activities, simplify compliance audits, discover and troubleshoot security and operational issues. Provide visibility into user and resource activities, and track and automatically respond to security threats within your AWS infrastructure. For example, you can utilize CloudTrail to automatically respond to security vulnerabilities. You can create a workflow to add a specific policy to an S3 bucket, when CloudTrail finds API call that made the bucket public. You track many account activities, including actions taken through the AWS Management Console, AWS SDKs, and command-line tools.

You can review logs using CloudTrail event history. Have the reports delivered to S3 buckets or send reports to CloudWatch logs and events for more granular monitoring of AWS resources. You can view, filter, and download account activities for the most recent 90 days for free. You can also, set up a trail that delivers a copy of management events in every region free of charge. However, the data is sent to S3 so you will be charged for storage usage. Data events, which are operations performed on, or within the resources itself also have very small charges. AWS CloudTrail is an invaluable resource in simplifying events security analysis and troubleshooting for your AWS cloud IT infrastructure.

CloudWatch

But now you need to actively monitor it and collect metrics and react to any events. Unfortunately, you can't be up 24/7 monitoring and neither can your team. Thankfully, Amazon CloudWatch is a monitoring and management system built for developers, system administrators, site-reliability engineers, and IT managers.

Natively integrated with over 70 AWS services, CloudWatch helps you gain system-wide visibility into resource utilization, application performance and operational health. It collects monitoring and operational data as logs, metrics and events to provide insight into your application performances. You can collect and track metrics in real time or have it send off notifications when an event occurs. You can even set up CloudWatch alarms to automatically make changes using pre-defined triggers so you don't have to lift a finger to fix common issues. CloudWatch employs a pay-as-you-go model so you only pay for what you use with no up-front commitment. Keep tabs on your applications hosted on AWS Cloud with Amazon CloudWatch so you and your teammates can get a good night sleep instead of holding round-the-clock monitoring vigils.

Study break: Reviewing management tools

In this chapter, we went over three of the major management tools in AWS:

- AWS CloudFormation,
- AWS CloudTrail, and
- Amazon CloudWatch.

Let's quickly review all of them to make sure we've got the fundamental concepts down before moving on.

AWS CloudFormation allows you to create a recipe for spinning up identical setups for a collection of resources and services for your IT infrastructure. It's free to use and you only pay for the resources you utilize by building a project on CloudFormation. It utilizes infrastructure as code, and you can deploy IT infrastructure based on a text file filled with code that specifies configurations for all of your services and resources. Once that's created, CloudFormation does the actual configurations and deployment for you. You can continue to build out your resources without having to worry about human error and configurations.

AWS CloudTrail can log monitor account activities, provide event history of account activities, simplify compliance audits, discover and troubleshoot security and operational issues, provide visibility into user and resource activities and track and automatically respond to security threats within your AWS infrastructure. And in that show, it's an event tracker and security analysis tool that helps keep your AWS cloud infrastructure compliant and secure.

Amazon CloudWatch helps you gain system wide visibility into resource utilization, application performance and operational health. It collects monitoring and operational data as logs, metrics and events and provides insight into your application performance. You can even set up CloudWatch alarms to automatically make changes using predefined triggers to automatically solve common issues. It's integrated with nearly 70 AWS services helping your team keep comprehensive monitoring data 24 seven.

Now you might be thinking CloudTrail, CloudWatch, what's the difference?

- AWS cloudTrail audits logs.
- Amazon CloudWatch monitors and can react to changes.
- Need access logs because someone did something they shouldn't have, CloudTrail.
- Need to know how much CPU on EC2 instance is using, CloudWatch.

Imagine a detective trailing a trail of footprints for CloudTrail. CloudWatch is watching or monitoring to make sure your resources are functioning as they should be. AWS's management tools help you build and manage your AWS cloud infrastructure. If any of the services we talked about needs a bit of a review, feel free to go check out the videos again.

Study break: Exam tips and resources

Let's quickly go over the services and concepts we'll be going over in this video. We'll be reviewing compute, storage, database, network and content delivery and management tools and services. And we'll also be reviewing concepts like infrastructure as code, deploying on the AWS cloud and availability zones and regions. As mentioned, we covered a lot of services in this course and many of their names are pretty confusing. I'm going to help you create a study cheat sheet so you can refer to it as you study for the exam.

Having a couple words that describe each service will make it easier for you to jog your memory when you're a little unsure about what the specific service does. Here we go.

Let's begin with compute services.

- Amazon EC2 or elastic compute cloud is a virtual server.
- Amazon Elastic Beanstalk helps you automatically grow your applications to meet demands, like Jack's beanstalk growing and growing.
- Elastic Load Balancing balances incoming traffic loads.
- AWS Lambda allows you to run serverless code.
- AWS Lightsail provides preconfigured virtual servers.

Let's move on to storage services.

- Amazon S3 or simple storage service provides object storage. Think of objects like individual files.
- Amazon Elastic Block Store provides block storage. Think of them like external hard-drives you'd attach to your computer. Block storage is, quote unquote, opposite of object storage like a zip file instead of a single image file.
- Amazon Snowball transfers huge amounts of data to AWS with a physical storage device.
- Amazon Storage Gateway provides gateways to connect on-premises resources with the cloud. Next, let's go over the database services.
- Amazon Dynamo DB is a non-relational or no SQL database.
- Amazon RDS or relational database service is a relational database.

- Amazon Aurora is a relational database and can be run on amazon RDS.
- Amazon Redshift is a data warehouse for a lot of data.

Now, to the network and content delivery services.

- Amazon VPC or virtual private cloud is a virtual network. Think of it as your corner of the AWS cloud.
- Amazon CloudFront helps you have speedy websites using edge locations.
- Amazon Route 53 routes domains to services and IP addresses.

Finally, let's go over the AWS management tools.

- Amazon CloudFormation helps you create templates to form cloud services.
- AWS CloudTrail helps you track trails of action, think of audit logs.
- Amazon CloudWatch monitors or watches your AWS cloud instance for you.

Just as a note, sometimes other services come up in the exam. However, if you know what these core services are and what they do, then you'll be able to filter through the ones you're unsure of to come to the correct answer. Thankfully it's a process of elimination because the exam is all multiple choice. So don't get distracted by a service you've never heard about and utilize your process of elimination skills.

Next, let's discuss infrastructure as code. The concept of infrastructure as code is that you can write code that describes the configurations for specific AWS cloud services and they can be deployed for you by AWS. It helps speed up the deployment process and removes the risk of human error when spinning up new resources.
Some AWS cloud services that utilize infrastructure as code are:

- Elastic Beanstalk,
- AWS Lambda, and
- AWS CloudFormation.

Some ways to deploy and manage resources on the AWS cloud are by utilizing AWS Management Console with AWS Command-Line Interface or CLI, and with AWS Software Deployment Kits, or SDKs.

Finally, let's review AWS as global infrastructure. AWS has data centers around the world called availability zones. Each availability zone is independent from each other in network and power source and there are currently almost six dozen availability zones or AZs around the world.

A region is made up of two or more availability zones and there are currently two dozen AWS regions around the world. You should strive to create a highly available, resilient and redundant IT infrastructure by replicating your AWS cloud resources across multiple availability zones and potentially, even regions.

AWS

Introduction to AWS for Non-Engineers

Module 4

Billing and Pricing

Billing and Pricing domain

There are four domains in the AWS Certified Cloud Practitioner exam. They are

1. cloud concepts,
2. security,
3. technology, and
4. billing and pricing.

The four courses in the Introduction to AWS for Non-Engineers series follow these four domains. This course that you're watching now covers the fourth domain, billing and pricing. The billing and pricing domain makes up the smallest portion of the certification exam.

For this domain, you will need to be able to do a few things. First off, you need to be able to compare and contrast the various pricing models for AWS. Second, you need to recognize the various account structures in relation to the AWS billing and pricing. Finally, you need to identify resources available for billing support.

There are four different support plans available with fees separate from monthly usage fees. The monthly support plan fees range from free all the way to starting at $15,000 a month depending on how much support your organization requires from AWS.

There are also different types of charges that occur when utilizing the AWS Cloud, such as compute, storage and data out. And there are ways to save money when you're a multi-account organization with multiple AWS Cloud accounts under one roof called consolidated billing. Even though the billing and pricing domain is the smallest section by far in the exam, the questions pertaining to this domain are fairly nitty-gritty. So we want to make sure you are well prepared. Let's get started.

Billing concepts

You've decided to host your static website on AWS and registered a domain through Route 53, AWS's domain name system, and hosted your website on S3, a storage service. You've now got bills to pay for your AWS use, what do you do? Log into your AWS admin console and check out the Billing and Cost Management Dashboard.

The AWS Billing and Cost Management Dashboard allows you to estimate and plan your AWS costs. Through a service called consolidated billing, you can simplify your accounting if you have multiple AWS accounts. You can also receive alerts for service usage thresholds, which could help keep you from spending more money than anticipated.

You can utilize a feature called Cost Explorer to view your costs as graphs, filter results by values like availability zone, AWS Services, EC2 instance types, region, usage types, and much more. You can also see a forecast of potential costs based on historical usage data. You can even have AWS generate billing reports with a breakdown of your costs by the hour or month, by product or by tags for your organization's billing needs. As you begin to deep dive into AWS and start testing out and utilizing its services, the Billing and Cost Management Dashboard will become a very important ally in both making sure your services don't get turned off for any billing related reasons, and to keep tabs on costs.

Types of charges

One of the biggest alerts of cloud computing platforms is the pay-as-you-go model for resources. Instead of having the huge upfront cost of buying physical service and setting up a data center that has to be maintained. Cloud computing platforms like AWS only charge for resources as you use them. You can easily scale your resources up or down to suit your businesses demands and only be billed for the resources you consume.

The three fundamental drivers of costs with AWS are

- compute,
- storage, and
- outbound data transfer.

For compute resources, you pay hourly from the time you launch a resource until the time you terminate it. Think about virtual servers you would pay for the amount of time the server is up and running.

For data stores or transfer, you typically pay per gigabyte. An example would be storing profile images uploaded to your app for a social media application on the Cloud. You'll pay more if you have more users uploading images because you're using more space. In most cases, there are no charges for inbound data transfers or data transfers between different AWS services within the same region. This could be your social media application, saving the profile images to the storage service, once the user clicks upload inside the app. Different services use different pricing models. For example, S3 AWS's storage service utilizes different prices by storage types. The most basic S3 storage type is the S3 Standard Storage. This storage type is the most expensive storage type because it boasts high availability and extremely low chances of data corruption. Even better, you can retrieve the data immediately. Compare this with S3 Glacier Storage, which is substantially cheaper, buffered a much lower price you must wait longer for the data to be retrieved and the availability is not as high. As a result the S3 Standard Storage is well suited for objects you will use often, like images live on a website, whereas S3 Glacier Storage will be better suited for backups of data for safekeeping. While S3 has different prices by storage types Aurora AWS has database service charges for storage and data input, and output. It's imperative even while using the AWS Free Tier to find out how AWS charges for service usage, so you aren't hit with surprise bills at the end of the month.

Consolidated billing

When a company embraces the cloud and all of the infrastructure engineering teams happily pounce onto AWS and begin building, a single company may end up with multiple AWS accounts. Maybe the development team wants to split up their production and test environments so there won't be any accidental deployments. Or maybe the marketing team wants their own instance for the company website. Other times it just might make sense for different projects to open different accounts to keep their billing responsibilities apparent. Whatever the reason may be, having multiple AWS accounts is likely a huge headache for the accounting team, when they need to keep track of all the accounts and numbers.

To make their lives easier, AWS introduces the consolidated billing option. As the name suggests, it allows an organization to create a payer account that views and pays combined billing charges for all linked accounts. It is strictly an accounting and billing account and can't use any other services. However, it's still an independent account, the account cannot deploy services into the linked accounts. Using consolidated billing comes with the great perk of all of the resource usage being considered part of one large organization. So even though they may be in separate accounts, the organization may be eligible for volume discounts for the combined usage. For busy organizations with many separate AWS accounts, consolidated billing is a must use for the accounting department to easily review and pay bills as well as take advantage of potential volume pricing discounts. Best of all, it's free!

Cost calculators

One of the most common questions people have about cloud services is how much will this cost? Fortunately, AWS provides a few tools to help estimate your costs. You'll be able to approximate your savings by moving from on-premises infrastructure to AWS cloud, as well as your monthly AWS bill. When considering whether it makes financial sense to keep your on-premises IT infrastructure or to move it over to AWS cloud, you can plug your specific requirements in the AWS Total Cost of Ownership Calculator.

You can get detailed reports of estimated cost savings by modifying some assumptions to match your infrastructure. The total cost of ownership, also known as TCO, can be reduced by moving to the cloud, because you no longer need to invest in large capital expenditures by purchasing all the hardware and infrastructure up front. Rather, you will utilizing a pay-as-you-go model that allows you to invest capital, both financial and human, only when your business has a need. You can visit the AWS TCO Calculator by searching in your favorite search engine or by going directly to awstcocalculator.com.

Another calculator is the AWS Pricing Calculator, which used to be the simple monthly calculator. You can find this one at calculator.aws. And it helps you estimate the cost of a cloud architecture solution you're looking to build. You would add services and configurations to the calculator, which would provide a report with estimated totals per service, service group, and total infrastructure. You can utilize this service to compare service costs per region, reduce your EC2 spend, find the right EC2 instance for your needs, or estimate your overall AWS cloud spend. Cost Explorer is an AWS tool that enables you to view and analyze your cost in usage, forecast how much you'd likely spend, or get recommendations for what reserved instances to purchase to minimize costs. The Cost Explorer is part of the Billing and Cost Management Console dashboard. The pay-as-you-go model is one of the largest allures of moving your IT infrastructure into the cloud. And tools like the TCO Calculator and Pricing Calculator help you get a better idea about the potential costs for moving to the cloud.

AWS Free Tier

Are you excited to try out AWS Cloud and play with the services, but a little nervous about the potential costs associated with using it? Fear not, the AWS Free Tier is here for you. The AWS Free Tier allows potential customers to test out and become comfortable with many services offered by AWS Cloud for free. Most of the offers expire after 12 months, at which point you'll be charged for the services you consume at regular rates. As you near the expiration date of the Free Tier, you'll get a notification from AWS. You are then responsible for manually turning down or deleting the services for which you don't want to be charged.

You can find the Free Tier offerings by going to AWS.Amazon.com/free. Here, you'll notice that AWS has over 60 services available for use separated into three different types of offers. The three types of offers in AWS Free Tier are always free, 12 months free, and trials. First option is always free, which, as you might guess, shows services that are always free to use up to a certain point. There are generally usage limits after which point you must pay to use these services. The second option is 12 months free. These services are free for use, also generally up to a certain usage limit, for 12 months after your initial sign-up date. Some of the common uses limitations are use time, number of requests, amount of storage, number of characters, and number of actions per month. Once you go above a usage limit, you'll be charged for the service you use, even if you're still within the first 12 months. The final type of Free Tier offerings are trials. Most of the trials are for less than 12 months and have stricter usage limits. The common limitations are use time, use of space or number of requests. The AWS Free Tier is a great opportunity to test out and learn about many of the core services that power the AWS Cloud, and I highly recommend that you go take advantage of the different offers.

Study Break: Billing and Pricing domain

Welcome to the billing and pricing study break. Let's review some of the concepts we learned about that will be good to know for the AWS Certified Cloud Practitioner Exam.

The AWS billing and cost management console is the go to a place to plan your AWS spending, simplify your accounts or consolidated billing and receive alerts for service uses this thresholds.

Cost explorer lets you get granular information about your AWS usage and generate billing reports with a breakdown of costs and usage.

The AWS billing and cost management console is an important part of making sure your AWS bills are expected and paid.

Additionally, the AWS pricing calculator and the AWS total cost of ownership calculator help you to evaluate how much running or migrating your it resources onto AWS Cloud could cost.

We just covered how to manage your billing. Now, let's cover the charges options provided by AWS.

The pay-as-you-go model of Cloud computing includes the different ways AWS charges you for your AWS Cloud resource usage.

The three fundamental ways you can be charged are compute, storage and outbound data transfer. Different services have different ways of charging and pricing. And in general, the more you do something compute store or transfer data, the cheaper per unit the action becomes. There are more ways that AWS services charge its users, but these are the three most common charges. You should always check the way service usage is charged before spinning up any resources in AWS Cloud that you have never used before, so that you aren't hit with surprised bills at the end of the month.

All of these charges, you may wonder, is there any way that I can save money? Yes, there is. Consolidated billing is one way of saving money, while making the administrative work of managing multiple AWS accounts within one organization much easier. Instead of manually logging into each account owned by an organization to check usage and pay bills. The organization can create a billing only payer account that views and pays combined billing charges for all linked AWS accounts. One huge advantage of consolidated billing is that all resources used within the linked accounts are considered part of one large organization. This means that the organization as a whole maybe eligible for volume discounts for the combined usage instead of by account basis.

Basic

Support is something most organizations need to a varying degree. To address this need, AWS offers four types of support plans for all users and organizations on its platform, depending on their budget, level of engagement with AWS, and support requirements. The cost can range from free all the way up to starting at $15,000 a month with different kinds of resources and support at every level.

It's important to keep in mind that the cost of the plan does not include usage costs, so this is the monthly cost to have the support option in case you need AWS's technical support. In addition, you will be paying usage fees separately. The cheapest option is the Basic Support Plan which is absolutely free. This level of support is perfect for users learning about AWS who might be spending some time testing the services and functions out. It goes very well with the AWS Free Tier, which is a 12-month period after you create your AWS account when you can try out many of the core services for free. This support plan offers no tech support, but does provide access to the AWS community forums where you can ask technical questions to other AWS users and engineers. Sometimes engineers working for AWS may step in and provide guidance or resources. Their customer service is limited to account and billing questions. With the Basic Support Plan, you have access to the seven core Trusted Advisor checks and guidance to help provision your IT resources into AWS Cloud using best practices. You also get a personalized view of the health of AWS Services and receive alerts when your resources are impacted through AWS Personal Health Dashboard. Since you won't have access to technical support from AWS, the Basic Support Plan is great for organizations or people testing out AWS without any mission-critical resources on the platform.

Developer

If you are in need of more support than the basic plan can offer, the next step up is the Developer Plan. The Developer Plan starts at $29 a month and scales with use. Scaling is the concept that an infrastructure grows and shrinks with your projects or needs. In this case, the plan starts at $29 a month but the ultimate monthly bill scales up with different bells and whistles added to fit your specific needs. AWS will charge you either $29 a month, or 3% of your AWS use costs, whichever is larger. This Developer Plan Support is perfect for people and organizations experimenting with AWS at a higher intensity than those using the Basic Level Support Plan, and may require a bit more technical assistance. One person in the organization is specified as a primary contact and they can ask technical questions through the support portal. This person can open an unlimited number of cases, and technicians will respond during business hours via email.

The SLA, or service-level agreement, defines the amount of time it will take before a technician will respond to a support case. For the Developer Support Plan, AWS commits to a technician following up within 12 hours for an impaired system or 24 hours for general guidance. Those typically on a Developer Plan will be testing out features and potentially deploying prototypes and resources to see if they are good fits for their organization's goals. Because they are digging deeper into how the AWS Services work, assessing them for viability within their IT infrastructure, having technical support is important. However, since they are not yet fully committed to AWS, they don't want to pay the higher fees associated with the higher level support plans. One thing to keep in mind is that the technical support is not anywhere near immediate, so this plan is not ideal for production use of resources where service failures could have severe business impact.

Business

The third Support tier is the Business Support Plan, which is an ideal plan for those who use AWS in production. The fees associated with the Support Plan begin at $100 a month, and scale up with use. AWS will charge you the higher value of 100 a month, or anywhere from 3 to 10% of your monthly AWS bill when you select this Support Plan. You can find out more by accessing aws.amazon.com/premiumsupport/plans.

An unlimited numbers of users in your organization, called contacts, can open an unlimited number of Technical Support cases at the Support Center. You also have access to the AWS Support API for support case automation. With Support API access, your developer team can retrieve detailed information about support operations and data types in JSON format. If you are subscribed to the Business Support Plan, you have access to the full suite of AWS Trusted Advisor checks, instead of just the seven core checks that come with the basic and developer plan. Trusted Advisor checks, whether basic or full suite, help you optimize your IT infrastructure hosted on the AWS Cloud. You can also retrieve lists of checks, check results, and refresh status of checks using AWS Support API mentioned earlier.

For an additional fee, you also get access to Infrastructure Event Management, which helps you plan for large-scale events, ranging from mobile app launches to IT infrastructure migrations into AWS Cloud. AWS provides planning assistance and real-time support during your event so you can proceed with confidence. The Business Support Plan also provides support for common third-party application stack components, operating systems, and platforms. AWS's support team will provide guidance, configuration support, and troubleshooting of AWS interoperability with many other third-party software. SLA for tickets in the Business Support Plan is 24/7 support via phone, email, and chat, and they strive to provide one-hour response time to your urgent support cases when a production system is down. The Developer Support Plan only offers email and Support Center support, whereas the Business Support Plan allows you to contact them via phone and chat as well. The Business Level Support Plan is the best bang for buck in terms of cost versus the level of support provided compared with other Support plans.

Enterprise

The fourth support tier is the Enterprise Support Plan, which is an ideal plan for those with mission-critical use of AWS. It comes with 24/7 technical support via email, chat and phone and has an SLA of 15 minutes for business-critical support cases with priority. Your organization can have unlimited number of contacts who can open an unlimited number of support cases. Like the Business Support Plan, Enterprise Support Plan comes with a full set of Trusted Advisor best practice checks, AWS Support API access, and third-party software support.

The Enterprise Support Plan is the only tier that provides Technical Account Manager, or TAM, and Support Concierge access. TAMs provide proactive best practices guidance which helps you develop and run your AWS infrastructure efficiently. This includes proactively monitoring your infrastructure and helping you optimize it. Support Concierge provides account and billing analysis to help you cut service fees. The support plan also provides various proactive programs like Infrastructure Event Management, which was for an additional fee with the Business Support Plan, but is complimentary with the Enterprise Support Plan. They also provide Well-Architected Reviews, which are detailed reviews of your architecture to guide you on how to best design your systems. They also provide Architecture Support to help you better align your infrastructure with AWS. And operations support which provides reviews of your operations and provide advice for optimization.

On top of this, AWS provides training by way of self-paced online labs provided through an AWS training provider. For all these perks and speedy service, you are looking at price tags starting at $15,000 a month. With the Enterprise Support Plan, AWS charges you for the higher of $15,000 a month or three to 10% of your monthly AWS usage bill. The Enterprise Plan is great for big organizations with mission-critical use of AWS who cannot afford to have long-standing downtime. They would also need to have fairly big wallets too, as the starting price of $15,000 a month does not include the usage charges for the AWS service being utilized.

Which one's best for you?

Which support plan should you choose for the AWS Certified Cloud Practitioner exam? If you're going to be taking the AWS Certified Cloud Practitioner exam, knowing what type of support plan is best for a sample organization's needs and budget is a question that comes up a lot. As we just learned, support plan prices can very drastically, from free to over $15,000 a month. And the type and speed of support also vary from tier to tier. When considering exam questions about appropriate support plans, you need to balance the support needs with the potential funding available.

Let's try a sample question: Company Y is a mid-sized company looking to migrate their IT infrastructure into the cloud. They are shopping around for the right fit and testing different cloud computing platforms. They have a deadline in choosing the platform, so they would like to test things out efficiently, but also without spending too much money. Which support plan should they choose? Take a stab at this question. Should they choose Basic, Developer, Business or Enterprise Support Plan?

The answer will be Developer Support Plan. They are just in a testing phase and don't want to spend a lot of money, but they also don't want to spend too much time poking around and figuring things out, so they would like some level of technical support for questions. The Developer Support Plan gives you 12 to 24 hour SLA for one technical contact to open as many tickets as they want for just $29 a month. The Basic Plan only allows you to ask questions at the support forum where you may or may not get a response, so it would not be a very efficient way to evaluate the platform for corporate use. Of course, Business and Enterprise Support Plans provide a lot of support but also come with heftier price tags. If they are just testing the services out, there's really no need for a full-blown business account, much less an enterprise account. What do you think, how about your company? Can you identify a support plan that would fit its current needs for IT infrastructure?

Study break: Reviewing AWS support plans

For this study break, let's review the four different support plans available for AWS Cloud. Questions about the AWS Support Plans appear in the AWS Certified Cloud Practitioner exam. So it's important to understand the differences and similarities between the four options, if you're looking into taking the exam.

Organizations can choose a support plan based on their budget, level of engagement with AWS, and support requirements. One important thing to remember when considering support plans in real use case scenarios, is that it is a separate charge from your usage cost.

The four different support plans are Basic, Developer, Business and Enterprise Support Plans.

The Basic Support Plan has no monthly fees and is a great way for both organizations and users to test out and learn about the AWS Cloud and to evaluate the different services and functions. With this plan, you will not be able to receive any tech support aside from access to the AWS Community Forums, or you can pose questions for other users. You can receive customer service for account and billing questions directly from AWS. The Basic Support Plan is best for organizations without any mission critical resources hosted on AWS Cloud.

The Developer Support Plans starts at $29 a month and scales with use. You will be charged either $29 a month or 3% of your AWS monthly usage cost, whichever is larger. So if 3% of your monthly AWS cost is less than $29, you'll receive a flat $29 bill. If it's larger than $29, you'll pay the larger fee. This plan is great for organizations that are experimenting with AWS in a more serious way for potential mission critical usage, and require more technical assistance than what the community forum can offer with the Basic Support Plan. However, the technical support is not immediate, so this plan is not ideal for production or mission critical use of AWS, where service failures could lead to business disruption.

The Business Support Plan starts at $100 a month and scales with use. As with the Developer Support Plan, AWS will charge you the higher value of either $100 a month, or between 3% to 10% of your monthly AWS bill. You can receive 24/7 tech support and for a fee could receive additional services like infrastructure event management. The Billing Support Plan is often considered the best bang for buck in terms of monthly cost versus the level of support compared with other support plans.

Finally, the Enterprise Support Plan. It comes with a hefty price tag starting at $15,000 a month and scaling with use. As with the Business Support Plan AWS charges the higher of the static fee, or 3% to 10% of monthly AWS usage bill This plan comes with all the bells and whistles, including a technical account manager, support concierge, proactive programs, well architected reviews and training, along with 24/7 tech support with a service level agreement or SLA of 15 minutes for emergencies. The Enterprise Support Plan is great for large organizations with substantial mission critical usage of AWS that cannot afford to have long standing downtime of their infrastructure. Remember, the support plan cost of $15,000 and up is an addition to the AWS usage bill. Aside from the costs and AWS use cases, there are more nitty gritty features and options with each support plan. Some options include number of contacts that can open support cases, and what the service level agreements or SLAs are for number of tickets.

Most of the questions about support plans are evaluating your ability to pick an appropriate support plan.

Study break: Exam tips and resources

Of the four domains in the AWS Certified Cloud Practitioner Exam, the Billing and Pricing domain has the smallest amount of content at 12% of the exam. However, the multiple choice questions on the exam require you to know the concepts inside and out. Especially important is the ability to compare and contrast the different support plans.

The first thing you need to be able to do is compare and contrast the various pricing models for AWS. The questions could ask about the different ways AWS charges for resource usage, with the most fundamental ones being compute, storage, and data transfer out. It may also require you to realize that for many AWS services, the more you do something, such as storage or compute, the cheaper per unit the action becomes. So, transferring 50 gigabytes of data may be cheaper per kilobyte of transfer than transferring just 400 megabytes of data. Another component of this domain is to recognize the various account structures in relation to AWS billing and pricing.

Questions about this section could be asking about consolidated billing, which helps lower organizational cost as a whole by creating a billing-only account that links all AWS accounts together within the organization. By doing so, the organization may be eligible for volume discounts by combining their resource usage from all the accounts.

The questions could also ask about different support plans and ask you to identify the most suitable support plan for a certain situation. There are four support plans available from AWS ranging in monthly fees from free to starting at $15,000 a month. The monthly fees do not include the monthly AWS usage costs, which are billed separately. The basic support plan is free and goes very well with the AWS free tier offer, which is 12 months of free service usage offered for new customers. And for the three plans that have monthly fees, AWS bills you for the higher of the monthly flat fee or somewhere between three to 10% of monthly AWS usage.

Are you thinking, "well, which is more expensive, "the business support plan or developer support plan?"

A silly memory aid that I came up with to memorize the four support plans in order of monthly cost was BDBE.

B for Basic, D for Developer, B for Business, and E for Enterprise.

The monthly prices and features provided go up in that order. Finally, you need to be able to identify resources available for billing support. These could be in the form of white papers, knowledge bases, contacting AWS Billing Support, or utilizing calculators like the AWS Cost Explorer, AWS Total Cost of Ownership Calculator, or the AWS Simple Monthly Calculator to find out how much you can expect to pay by running your resources on the AWS cloud.

How do you feel about your support plan compare and contrast skills? Or the different ways that you can be charged by AWS? If you have a few minutes, I highly recommend that you pick out a service like Amazon EC2, and search around the official AWS website for ways they bill for service usage and find out how you can reduce the cost of running that service. The best way to digest information is by testing it out, and you'll learn a lot from going out and trying to find the information yourself for future use.

Next steps

Well, that was a lot of information in such a short amount of time. I'm so glad you stuck with me to the end. I hope you not only learned a few things, but enjoyed the process too. If you are interested in learning more about Amazon Web Services, and even potentially taking the AWS Certified Cloud Practitioner exam. The courses cover the four domains of the AWS Certified Cloud Practitioner exam, which are cloud concepts, security, technology, which we refer to as core services, and billing and pricing. If you have questions or want to learn more about cloud computing and potential careers that work with or in cloud computing, please come visit Cloud Newbies, a community of cloud newbies and seasoned pros, where we learn about cloud computing and study for certifications together. You can visit us at cloudnewbies.com. If you're looking for a resource website, while you're beginning your research into Amazon Web Services, you can visit me at awsnewbies.com, where I introduce cloud computing and AWS in a jargon-free way. Thanks again for watching and I hope to see you again in one of my other courses or resources. Good luck!