**AWS**

With AWS, you have the ability to only pay for what you use. For example, you could use Amazon EC2 to launch as many or as few virtual servers as you need.

Cloud computing is the on-demand delivery of IT resources over the internet with pay-as-you-go pricing.

**How is AWS able to offer pay-as-you-go?**

AWS separates different virtual machines from each other on the same hardware. So for example, my company and John’s company could be sharing the exact same physical server owned by AWS, but our virtual machines would still be completely separated. This ensures that servers are put to their highest and best use, allowing cheaper server costs to us as users than what we’d typically have to pay for to manage our own data server (unless my company becomes as big as Amazon).

--------------------------------------------------------------------------------------------------------------------------------------

**AWS EC2**

You can use Windows or Linux OS with EC2.

Any software, databases, web apps, scripts, and more can run on the EC2 instance

There are a variety of tasks that can be done, so different EC2 instances should be used for specific jobs. Each instance type is grouped under an instance family, similar to how employees of a big company are run under teams (marketing, sales, engineering, HR, etc).

EC2 instance families:

* General Purpose: It can be used for diverse workloads, web servers, or code repositories.
* Compute Optimized: used for compute-intensive tasks like gaming servers or scientific modeling.
* Memory Optimized: Memory-intensive tasks.
* Accelerated Computing: Floating point number calculation, graphics processing, streaming data, or data pattern matching.
* Storage Optimized: High performance for locally stored data.

You can pay for it on demand (per hour or per second, depending on instance type and OS).

You can buy a savings plan (lower prices in exchange for a commitment of dollars per hour for a 1 or 3-year term). This can save up to 72% of AWS costs.

You can also have reserved instances. These are good when you know you will need a steady amount of resources that is unlikely to change.

You can also pay for spot instances. These can offer up to 90% off the on-demand cost, but AWS can stop your instance at any time (with a 2-minute warning to save your work). This is best used when your workflow can be interrupted, such as batch workloads.

The final payment option is to pay for a dedicated host. No one else will share the host, and these are usually for compliance requirements.

--------------------------------------------------------------------------------------------------------------------------------------

* If you have multiple EC2 instances all running the same task, a load balancer makes sure that each EC2 instance has an even distribution of workload between them.
* **Elastic Load Balancing** is the AWS service that automatically distributes incoming application traffic across multiple resources, like EC2 instances.

--------------------------------------------------------------------------------------------------------------------------------------

If the architecture is tightly coupled, any errors with one application can cause cascading problems with any other applications it’s tightly coupled with. However, if it’s loosely coupled, errors in one app may not necessarily affect the other app.

**To create a loosely coupled architecture:** You can introduce a message queue between applications as a sort of buffer. The message queue takes in messages from App A and delivers them to App B when App B is ready. If App B fails, App A does not experience any disruption and the messages will remain in the message queue until App B is back and ready to handle them again.

**2 AWS services that act as message queues**

AWS SQS (Simple Queue Service): can send, store, and receive messages at any volume

A payload is the data contained within a message.

SNS (Simple Notification Service): It can send notifications and messages to “subscribers”, which can be phone numbers, email addresses, web servers, AWS Lambda Functions, and more.

**AWS Lambda :** Lambda is a serverless computing platform that can run your script after an event is triggered. It is also meant to run code for under 15 minutes, and is more suited for quick processing.

* You also won’t have to worry about scaling the environment either.

If you need access to the underlying environment, you need a container orchestration tool, such as AWS Elastic Container Service (ECS) or Elastic Kubernetes Service (EKS).

A **container** is a package for your code. When you use a Docker container, it will be running on the EC2 instance.

If you want to host a traditional application and want access to the underlying OS, you’ll want to use AWS EC2.

If you are going to host short-running functions and don’t need access to the underlying environment at all, you’ll want to use AWS Lambda.

If you’d like to run Docker container-based workloads on AWS, you’ll need to choose AWS ECS or EKS, then you can run them on EC2 instances that you manage, or in a serverless environment like AWS Fargate that is managed for you.

--------------------------------------------------------------------------------------------------------------------------------------

**AWS Global Infrastructure:**

* An **availability zone** is a group of locations where AWS data centers are. They are spread out far enough so as not to have all of them fail at once in a disaster, but close enough to still have low-latency.
* You’ll want to run your application/instance on at least 2 availability zones. In case something happens to one, you’ll have another zone up and running.
* An **edge location** is a site somewhere in the world that **Amazon CloudFront** uses to store cached copies of your content closer to your customers. So if your data is in the Ohio Region but you have customers in China, you can cache a copy of your data somewhere close to China so your customers in China will receive your data faster.

**How do you work with AWS?**

* You can work with AWS resources using the Management Console, or the Command Line Interface (CLI), or Software Development Kits (SDK), or other tools.
* The Management Console is on your browser.
* You can create instances manually through the Management Console, but if you have to create multiple instances, you open yourself up to manual errors. The CLI can help automate these tasks.
* The SDKs allow you to access AWS using a programming language.

**AWS Elastic Beanstalk** helps you provision AWS EC2-based environments. You can assign your code and resource provision requirements to the Elastic Beanstalk service and it will create your environment for you.

* It can also save your environment configurations to use again at a later time.

**AWS CloudFormation** is an “infrastructure as code” tool that allows you to control and define many AWS resources using “CloudFormation Templates” in the form of JSON or YAML.

--------------------------------------------------------------------------------------------------------------------------------------

**AWS Networking**

Amazon VPC(**Virtual Private Cloud)** lets you create an isolated, virtual network that can either be public or private. It’s your own private network in AWS. You can deploy things like EC2 instances and your ELB in your VPC.

**Subnets** are chunks of IP addresses in your VPC that allow you to group resources together.

An **internet gateway** is like a door that lets public traffic into your VPC.

A **VPG (virtual private gateway)** into your VPC that only lets traffic in from specific sources you choose (an approved network), such as from your on-premise data center.

The problem with a virtual private gateway is you could still be subject to slowdowns when moving data across the internet. But with **AWS Direct Connect**, you can create the lowest latency and highest security dedicated fiber connection from your data center to AWS. This connects a physical line from your data center to your AWS VPC.

A **Network Access Control List (ACL)** controls what traffic is let into the subnet. The ACLs check traffic going in and out of the subnet. Only traffic that is approved beforehand is let in.

A security group in an instance-level ACL, and only a specific traffic type that is approved beforehand is let in (like HTTPS or OS traffic).

The Network ACL is like a Passport Control Traffic officer checking traffic in and out of the country, while a Security Group checks traffic going into a building (but doesn’t care about traffic going out).

The Security Group is also **statefull** (while the Network ACL is **stateless**), which means the Security Group can remember which packets are already cleared for entry and therefore does not need to check a packet’s validity if it’s already been accepted through the instance before. Meanwhile, the Network ACL checks a packet’s validity every single time it passes through a subnet, regardless if it’s already been accepted before.

**What is Amazon Route 53?**

Route 53 is Amazon’s DNS.

**What is DNS?**

The Domain Name System translates website names into IP addresses that computers can read.

**What’s the CDN?**

The Content Delivery Network helps deliver edge content to users based on their geolocation.

**AWS Databases**

**What is instance storage?**

Instance storage is storage attached to the physical hardware running your instance, and can store temporary data. However, it is deleted the moment your instance stops running. It is fine to use with temporary data you don’t actually need to keep.

**What is Amazon Elastic Block Store?**

AWS EBS can create virtual hard drives that you can attach to your EC2 instance for purposes of storing data.

EBS volumes come in different sizes and types. When you create an EBS volume, you can define the size, type, and configurations. Then you can attach to your EC2 instance.

**What are snapshots?**

Snapshots are incremental backups, which your EBS volume should be doing regularly to backup your data.

Incremental backups are different from **full backups.**Incremental backups only backup data that has changed since the last backup (similar to how Git only tracks changes and not the entire repo every time you commit).

**What is Amazon Simple Storage Service?**

AWS S3 allows you to store and retrieve an unlimited amount of data. You store objects (like files) in buckets (like file directories). The maximum object size is 5 TB. You can also version objects so there is always a copy of the last version (similar to how Git can revert to a previous change or restore something that was deleted in the current commit).

**What are the different tiers of S3?**

The first is **Amazon S3 Standard**. This is standard storage, and is meant for frequent retrieval of data.

You can also use Amazon S3 static website hosting, where you can store multiple html files and other web assets into a bucket, and then host it a static website.

The next tier is **Amazon S3 Standard-Infrequent Access (S3 Standard-IA)**. This is good for long-term storage, backups, and recovery files. It still allows rapid retrieval of data, but is best for data that isn’t accessed frequently. It has a lower storage price than S3 standard, but higher retrieval price. A variation of this is **Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)**, which only stores your data on one availability zone (for a cheaper price).

Another tier is **Amazon S3 Intelligent-Tiering**. For an additional monitoring and automation fee, it will automatically move data to different tiers based on retrieval. For example, if an object in S3 Standard hasn’t been accessed for 30 consecutive days, Intelligent-Tiering will move it to the Infrequent Access tier.

Another tier is Amazon S3 Glacier, which archives data. There are different retrieval tiers: **S3 Glacier Instant Retrieval**, **S3 Glacier Flexible Retrieval** (able to retrieve objects from a few minutes to a few hours), and **S3 Glacier Deep Archive** (12–48 hours to retrieve data).

There is also **Amazon S3 Outposts**, which allow you to create buckets on AWS Outputs (this is good for meeting local data residency requirements).

**What is S3 Lifecycle Management?**

Amazon s3 Lifecycle Management is where you can automatically move data to different tiers. For example, if you wanted data to be accessible in S3 Standard for 90 days, then moved to Standard-IA for another 30 days, and then moved to S3 Glacier after that, lifecycle policies can do this automatically.

**How does object storage work?**

In object storage, each object consists of: data, metadata, and a key.

The data is the data itself you are storing. The metadata is data about the data. And the key is a unique identifier for the object.

As a note, if you modify a file in block storage, only the pieces that are changed are update. But if you modify a file in object storage, the entire object is updated.

**What is Amazon Elastic File System?**

Amazon EFS allows you to have multiple instances accessing the data in EFS at the same time.

You should use “file storage” over block storage or object storage when you have a large amount and services and resources that need to access the same data at the same time.

**What is Amazon Relational Database Service?**

Amazon RDS supports all the major relational database management systems (RDBMS) like MySQL, PostgreSQL, and more. Amazon RDS also automatically handles things for you like backups, redundancy, and disaster recovery.

**What is Amazon Aurora?**

It’s Amazon’s most managed RDBMS option and offers MySQL and PostgreSQL, and is 1/10th the cost of commercial databases. It also has data replication and up to 15 read replicas (to scale performance and helps with high volume read traffic). Amazon Aurora also automatically creates continuous backups to Amazon S3. Amazon Aurora is best if your workloads require high availability.

**What is Amazon DynamoDB?**

AWS DynamoDB is a serverless, non-relational, NoSQL database. This is great for datasets that have variations from item to item. You can query this database through key-value pair relationships (think JSON or Python Dictionaries).

**What is a Data Warehouse good for?**

A Data Warehouse can be good for historical data which we will want to run Business Intelligence Analytics on. Historical data is set and will never change, making a Data Warehouse a good place to store it.

**What is Amazon Redshift?**

Redshift is Amazon’s Data Warehouse.

**What is Amazon Database Migration Service?**

Amazon DMS can migrate data from one source database to another target database (which can be the same type or different type).

If the database types are the same, this is known as a homogenous migration, and is a one-step (one-click) process.

If the source and target database types are different, this is known as a heterogeneous migration. You’ll need to use the AWS Schema Conversion Tool first to convert the source schema and code to match that of the target database. Then you can use DMS to migrate into the target database.

**What are other use cases for Amazon DMS?**

1. You can create a test database migration, so you can test applications against production data without affecting production users.
2. You can consolidate several databases into one database.
3. You can do continuous replication to send ongoing copies of your data to target sources (instead of a one-time migration).

**What is Amazon DocumentDB?**

DocumentDB (which has MongoDB compatibility) is great for content management, user profiles, and catalogs.

**What is Amazon Neptune?**

Amazon Neptune is a graph database.

**What is Amazon Managed Blockchain?**

It is an immutable, blockchain-based database.

**What is Amazon Quantum Ledger Database?**

Amazon QLDB is an immutable ledger database, where any entry can never be removed.

**What are some ways to accelerate your database performance?**

You can add caching layers onto your database to improve read times from milliseconds to microseconds. Amazon ElastiCache can provide these caching layers.

If you’re using DynamoDB, you can use DynamoDB Accelerator (DAX) which can dramatically improve read times for your data.

**AWS Security**

AWS has a shared responsibility model. AWS is responsible for the security **of** the cloud, and the user is responsible for the security **in** the cloud.

**What exactly are AWS and the user responsible for?**

AWS is responsible for the security of the hardware, network, and hypervisor (the software that creates virtual machines).

The user is responsible for the security of the operating system, your application, and your data.

**What is the AWS account root user?**

The root user has access and control of any resource in the account. It is important to also implement MFA (multi-factor authentication).

**What is AWS IAM?**

AWS Identity and Access Management (IAM) is a way to manage and control access to resources. You can create IAM users, and by default, will have no permissions.

You should follow the principle of least privilege: A user should be granted access only to what they need.

**What are IAM groups?**

IAM groups are groups of users with the same permissions (which are implemented via policies).

**What are roles in IAM?**

Roles have associated permissions that allow or deny specific actions (and can be assumed for temporary amounts of time). They are basically temporary permissions for users.

**What are AWS Organizations?**

AWS Organizations is a central location to manage multiple AWS accounts. It can also consolidate billing (using the primary account to pay for all accounts). You can also create hierarchical groupings of accounts. You will also have control over the AWS services and APIs that each account can access.

You can use Service Control Policies (SCPs) to specify the maximum permissions for accounts in the organization. SCPs can applied to Organizations Units (OUs) or individual member accounts.

IAM policies can be applied to IAM users, IAM groups, and IAM roles.

**What is AWS Artifact?**

AWS Artifact is a place where you can request documents proving that AWS is compliant with compliance.

You can also sign an **AWS Artifact Agreement** for certain compliance laws, such as HIPAA.

You can also access evidence of AWS security controls through **AWS Artifact Reports**.

**What is the Customer Compliance Center?**

The Customer Compliance Center is a place that contains resources to help you learn more about compliance.

**What is DDoS?**

DDoS stands for distributed denial of service, and is an attack on your infrastructure. It tries to overwhelm your system to the point it can no longer operate.

**What is a Slowloris attack?**

A Slowloris attack is a type of DDoS attack where the bad-acting client takes a very long time sending packets to the server. This is similar to how someone ordering a coffee shop and taking a very long time to order can create a long line of people waiting to give their order (essentially “clogging” the server). Because the Slowloris never finishes its request, the server will remain open, fill up their maximum concurrent connection pool, and eventually deny service from any other clients.

**How does AWS architecture help prevent DDoS attacks?**

Security groups only allow traffic which you specify. Security groups operate at the AWS network level (not at the EC2 instance level). If the traffic isn’t what you approved, it isn’t let in to clog the server.

**What is AWS Shield and what is AWS WAF?**

AWS WAF uses a web application firewall to filter incoming traffic with the signatures of bad actors (it uses a web access control list, or web ACL). It has ML abilities and can detect threats as they evolve.

AWS Shield provides DDoS protection. It is free if you don’t use the Advanced version.

**What is Encryption?**

Encryption secures your data in a way that only authorized parties can access it.

**What is AWS KMS?**

AWS Key Management Service manages your encryption keys.

**What is Amazon Inspector?**

Amazon Inspector helps improve security and compliance of your applications by running an automated security assessment against your infrastructure. It exposes your vulnerabilities.

It consists of 3 parts:

1. A network configuration reachability piece,
2. An Amazon agent, and
3. A security assessment service

**What is Amazon GuardDuty?**

Amazon GuardDuty analyzes your metadata to identify threats. It runs independently and does not affect performance of your workloads.

**AWS Monitoring and Analytics**

**What is monitoring?**

Monitoring means observing systems, collecting metrics, and then using that data to make decisions.

**What is Amazon CloudWatch?**

CloudWatch allows you to monitor your AWS infrastructure and your applications in real-time.

You can create a CloudWatch alarm to alert you once an event or metric is triggered. You can connect it to AWS SNS to send a message to a user to alert them as well.

You can create a CloudWatch dashboard to show us everything we want in one place.

**What is AWS CloudTrail?**

CloudTrail is an API auditing tool. Every request made to AWS gets logged in the CloudTrail engine (who made it, when, what was the IP address, and what was the response, was the request denied, etc).

You can also enable **CloudTrail Insights**, which can detect unusual API activities in your account.

**What is AWS Trusted Advisor?**

AWS Trusted Advisor is an automated advisor that evaluates your resources against 5 pillars:

1. Cost optimization
2. Performance
3. Security
4. Fault tolerance
5. Service limits

**AWS Pricing and Support**

**AWS Free Tier**

AWS Free Tier offers 3 ways to try out AWS:

1. Always free (only some services)
2. 12-months free
3. Trials (short-term free trials)

Here are some examples:

AWS Lambda allows 1 million free requests per month (and up to 3.2 million seconds of compute time… that’s 888 hours).

AWS S3 is free for 12 months for up to 5 GBs of storage.

AWS Lightsail offers a 1 month trial of up to 750 hours of usage.

A few more examples are AWS SageMaker, DynamoDB (25 GB of free storage per month), SNS, and more.

**AWS Support Plans**

There are 4 different types of support plans: Basic, Developer (email access to support), Business (phone access to support), and Enterprise (15-minute response time for mission-critical workloads, and dedicated Technical Account Manager (TAM)).

**What are the 5 pillars of the Well-Architected Framework?**

1. Operational excellence
2. Security
3. Reliability
4. Performance efficiency
5. Cost optimization

A TAM can also review your account and suggest changes based on the 5 pillars of the well-architected framework.

**What is AWS Marketplace?**

AWS Marketplace is a catalog that includes software from thousands of independent vendors that can run on AWS.

**AWS Migration and Innovation**

**What is AWS CAF?**

The AWS Cloud Adoption Framework is a framework to help users migrate on-premise data (or cloud data) into AWS.

The CAF has 6 perspectives:

1. Business
2. People
3. Governance
4. Platform
5. Security
6. Operations

**What are the 6 R’s of migration?**

Once you know exactly what you have in your existing environment, you decide which of the 6 R’s are best for your migration:

1. Rehosting (lift and shift) — Simply lifting your app/infrastructure and shifting it to AWS
2. Replatforming (lift, tinker, and shift) — You may make a few cloud optimizations first before lifting and shifting to the cloud
3. Retire — Some parts of your infrastructure may no longer be needed
4. Retain — Some parts of your infrastructure may need to be retained on-premise but not necessarily migrated to AWS (and can be retired eventually)
5. Repurchase — Abandon legacy software and move to something new on the cloud
6. Refactoring — You’ll need to write new code, you’ll need to add new features, and likely change the architecture of the code to make sure everything works once moved to AWS

**What is the AWS Snow Family?**

The AWS Snow Family is a collection of physical devices that help physically transport data into and out of AWS. For example, Amazon can ship you AWS Snowcone, you plug it into your data server or computer, copy all the data you’d like onto it, then ship it back to Amazon. Then, you will find all of your data copied into an S3 bucket, as an example.

**AWS Snowcone** is the smallest, holding up to 8 TB of storage.

**AWS Snowball** is the next biggest and comes in 2 versions: Storage-optimized (80 TB) and Compute-optimized (42 TB).

**AWS Snowmobile** is the biggest and is delivered via truck, holding up to 100 Petabytes (100,000 TB).

**AWS Cloud Wrap-Up**

What is the AWS Well-Architected Framework?

1. Operational excellence — Ensuring your architect delivers business value and continual improvement
2. Security — Protecting your data and systems with encryption
3. Reliability — Focuses on recovery planning from failure or disruption of infrastructure or AWS, as well as being able to scale horizontally
4. Performance efficiency — Choosing the right services (and the right types of a chosen service) for efficiency
5. Cost optimization — Optimizing for cost
6. Sustainability — Reducing energy consumption by minimizing the total resources needed an maximizing the benefits from the resources you are using

**What is the AWS Well-Architected Tool?**

The Well-Architected Tool is a tool you can run against your framework which can help check if your architecture meets the Well-Architected Framework standards (and can note areas of improvement).

**What are the 6 main benefits of using AWS?**

1. Trade fixed expense for variable expense (pay-as-you-go)
2. Benefit from massive economies of scale
3. Stop guessing capacity
4. Increase speed and agility
5. Stop spending money running data centers
6. Go global in minutes

**Additional Notes**

**What are shared controls?**

Shared controls are controls where both AWS and the user share responsibility, such as:

1. Patch management
2. Configuration management
3. Awareness and training

**Note about S3**

S3 cannot be scaled manually, it is instead scaled automatically. It can also provide unlimited storage for any type of data, but each individual object does have a maximum size of 5 TB. It is “infinitely scalable.”

**Why use Amazon ElastiCache?**

ElastiCache is a web service that makes it easy to deploy and manage an in-memory data store or cache in the cloud. It’s great for applications that need fast retrieval of data (querying a database is always slower than getting a copy of that data in the cache).

**What are the 3 AWS Cloud Computing Models?**

The 3 AWS Cloud Computing Models are:

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)

Note: Networking is a part of Infrastructure as a Service.

**What is an AWS-Managed Service?**

An AWS-managed service is a service where AWS is responsible for the operational and maintenance burdens of running them.

Examples are services like DynamoDB, Lambda, Elastic MapReduce, S3, and more.

IAM , VPC, and EC2 are NOT examples, since you are the one responsible for managing them.

**What is Infrastructure Event Management?**

Infrastructure Event Management is a short-term service you can get from AWS Support (which is included in the Enterprise Support plan and available for purchase in the Business Support plan) that helps provide architectural guidance for big traffic events (like product launches and advertising launches).