front	back
S3 object storage classes	- standard
S3 standard	Multi-AZ, single region
S3 intelligent tiering	Objects within the bucket are moved to infrequent access tier when not accessed for 30 days; when an object in IA is accessed, it is moved back to frequent
S3 standard IA	Good for infrequently accessed data
S3 one-zone IA	Good for infrequently accessed data when you can trade off cost for reduced availability
Glacier	Cold storage
Glacier deep archive	Cold storage
S3 lifecycle policies	Can transition objects from standard to IA to Glacier after a certain period (restrictions apply -- for instance, an object can't be transitioned to glacier less than
S3 lifecycle policies - minimum storag	- Standard: none
S3 versioning	With versioning enabled on a bucket, overwriting an object generates a version ID for the object; old versions are preserved.
S3 object lock	Available for all storage classes
S3 transfer acceleration	Use CloudFront to speed up transfer to/from S3 (there is a cost associated with this)
S3 events	Can be routed to:
S3 static websites	- enable web hosting
S3 security best practices	- block public access
EFS storage classes	- Standard
EFS performance mode	- general purpose (7K iops)
EFS throughput	- bursting: volume builds up crediets based on the filesystem size; credits allow bursting for limited time periods
Mounting EFS	- use /etc/fstab inside of linux VMs
EFS encryption	Encryption at rest supported via AWS-managed keys
Importing data to AWS	- Snowball
AWS Data Sync	Uses a super-efficient, purpose-built data transfer protocol that can run 10 times as fast as open source data transfer.
Snowball	Physical device shipped to your location; comes in 50TB and 80TB sizes (slightly less usable)
Snowmobile	100PB of storage capacity housed in a 45-foot long High Cube shipping container that measures 8 foot wide, 9.6 foot tall and has a curb weight of approxima
Storage Gateway	Hybrid cloud storage solution running on an on-prem VM or hardware appliance
Disaster recovery strategies	- Backup/restore
RTO	Recovery Time Objective - amount of time service can be offline
RPO	Recovery Point Objective
EC2 general-purpose instance types	Nitro-based:
EC2 Compute-optimized instance type	Nitro-based:
EC2 Memory-optimized instance types	Nitro-based:
EC2 Accelerated computing	Hardware acccelerators
EC2 Storage-optimized instance types	- I3: Intel + NVMe
Nitro	Underlying virtualization infrastructure for current-gen EC2 instances.
Graviton	Custom Arm-based processor designed to provide optimal price-performance ratio.
Inferentia	AWS custom silicon for deep learning.
EC2 instance lifecycle	INSTANCE LIFECYCLE DIAGRAM
Enhanced networking	Use Elastic Network Adapter (ENA) to support network speeds of up to 100 Gbps
Placement group	Placement groups influence the placement of a group of interdependent instances:
EBS optimized	EBS optimized instances deliver dedicated bandwidth to EBS.
EC2 user data	Small chunk of data (16KB max) that must be base64-encoded
EC2 burstable instance types	T2, T3, T3a, T4g
EC2 AMIs	- EBS-backed:
EC2 metadata	Instance metadata is data about your instance that you can use to configure or manage the running instance. Instance metadata is divided into categories, fc
EC2 instance store	An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer
EC2 pricing models	- On-Demand: expensive, no commitment
EC2 On-Demand	- most expensive
EC2 Reserved Instances	- up-front payment in exchange for lower prices
EC2 Spot instances	- pay market rates
EC2 Savings Plans	- optional up-front payment
EC2 Dedicated Instances	- physical EC2 server dedicated to your use
EC2 root volumes	- Instance store:
EBS	Elastic Block Store
EBS volume types	SSD:
EBS snapshots	point-in-time snapshots of your volumes to Amazon S3.
EBS encryption	Seamless encryption of EBS data volumes, boot volumes and snapshots, eliminating the need to build and manage a secure key management infrastructure
EBS: Data Lifecycle Manager for EBS	automated way to back up data stored on EBS volumes by ensuring that EBS snapshots are created and deleted on a custom schedule. No scripts or exterr
EBS elastic volumes	Elastic Volumes allows you to dynamically increase capacity, tune performance, and change the type of any new or existing current generation volume with r
S3 encryption	Encryption at rest:
ELB	An Elastic Load Balancer distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and Lambda functions.
EC2 autoscaling groups	A collection of EC2 instances treated as a logical group for purposes of scaling
Autoscaling policies	Types:
Autoscaling: simple scaling	The original scaling model for AWS Autoscaling groups
Autoscaling: step scaling	Scaling can be specified for *how much* a CloudWatch alarm is breached.
Autoscaling: target tracking	You set a target value for a metric (e.g. CPU load), and the ASG automatically scales up and down to try to maintain that target value.

Autoscaling: warm up	During a specified warm-up period, new instances are not counted toward the aggregated metrics of the group; this prevents excessive spin-up
Autoscaling: cool down	After a scale-up occurs, the ASG waits for a cooldown period to complete before any further scaling activities can start (only applies to simple scaling)
Autoscaling: notifications	Amazon EC2 Auto Scaling supports sending Amazon SNS notifications when the following events occur:
Autoscaling: launch configurations	A launch configuration that specifies things like:
Autoscaling: launch templates	A launch template is similar to a launch configuration, but it allows versioning; several versions can share some common configuration (e.g. the AMI), but differ in other configuration values (e.g. the instance type)
Security groups	A security group acts as a virtual firewall for your EC2 instances to control incoming and outgoing traffic.
Elastic IPs	a static public IPv4 address designed for dynamic cloud computing.
VPCs	A logically separated portion of the AWS cloud. Provides for:
Subnets: public	A public subnet is a subnet that's associated with a route table that has a route to an Internet gateway.
Internet gateway	An internet gateway is a horizontally scaled, redundant, and highly available VPC component that allows communication between your VPC and the internet.
NAT gateway	You can use a network address translation (NAT) gateway to enable instances in a private subnet to connect to the internet or other AWS services, but prevent the internet from initiating a connection with those instances.
NACLs	Network ACL
ELB types	- Application Load Balancer (ALB)
ELB: LCU	Load Balancer Capacity Units - used for billing by Application and Network load balancers
ELB: internal vs external	An internal ELB has only a private IP address and routes traffic within the VPC.
ELB: listener	ALBs use listeners -- a listener is a process that checks for connection requests, using the protocol and port that you configure.
ELB: health check	The ELB periodically makes requests to the targets to determine their health.
ELB: multi zone	Need to enable the AZ for the ELB, and you need to add targets in the AZ
Lambda	Serverless platform
Lambda: API Gateway	API gateway routes HTTP requests to Lambda functions
Lambda functions	Basic settings:
Lambda: supported languages	Java, Go, PowerShell, Node.js, C#, Python, and Ruby
Lambda: layers	A layer is a ZIP archive that contains libraries, a custom runtime, or other dependencies.
Lambda: database proxies	You can define an RDS proxy for your function
Lambda: VPCs	When you connect a function to a VPC, Lambda creates an ENI for each combination of security group and subnet in your function's VPC configuration
Lambda: permissions	The execution role grants it permission to access AWS services and resources
Lambda: invocation	Invocation can be asynchronous or synchronous
Lambda: autoscaling	Autoscaling accomodates an intial burst, followed by a gradual scale-up
EC2 reserved instance types	Offering classes:
EC2 capacity reservation	reserved instances that are AZ-specific come with a capacity reservation
RDS	Relational Database Service
RDS Database Types	- Aurora
RDS Multi-AZ	- a standby replica of the database is maintained in another AZ
RDS Read Replicas	- read replicas are read-only replicas that allow you to horizontally scale up a read-heavy application
RDS Backup	- RDS creates and saves automated backups of your DB instance during the backup window of your DB instance
RDS authentication	All RDS DB types support password authentication.
Aurora	MySQL and PostgreSQL-compatible relational database
Aurora serverless	Fully auto-scaled; you don't specify a number of instances
Aurora Global Database	Aurora database replicated across regions
S3 Security	- Bucket policy (what principals can do what to this bucket)
DynamoDB	Fully managed NoSQL key/value and document database
DynamoDB consistency	- Eventually consistent reads (default)
DynamoDB durability/availability	data is stored on SSDs and is automatically replicated across multiple AZs within an AWS region
DynamoDB capacity modes	On-demand capacity mode: charged for data reads/writes on your tables
DynamoDB pricing	charged for:
DynamoDB Accelerator (DAX)	fully managed, highly available, in-memory cache for DynamoDB
DynamoDB Streams	Captures a time-ordered sequence of item-level modifications in a DynamoDB table and stores this information in a log for up to 24 hours
DynamoDB backup/restore	- on-demand backups: full backups of tables; charged per GB of data stored
DynamoDB encryption	All customer data is encrypted at rest by default
DynamoDB keys/indexes	- primary key
DynamoDB primary keys	- simple primary key: uses a single attribute to identify an item (e.g. OrderID)
DynamoDB local secondary indexes	- must be specified at table creation
DynamoDB global secondary indexes	- used to query items across partition keys
DynamoDB item types	- strings
ElastiCache	fully managed key/value storage (faster than DynamoDB, but not durable)
ElastiCache - Memcached	- useful for transient data like cache and session store
ElastiCache - Redis	- redis is similar to memcache, but it has much richer queries and data types
Elasticache - Security	- runs in your VPC
Route53	A highly available and scalable DNS service. Provides three main functions:
Route53: Simple Routing	Standard DNS records with no special routing (like weighted or latency)
Route53: Weighted Routing	Allows you to associate multiple resources with a single domain name and choose how much traffic is routed to each resource
Route53: Latency-based Routing	Cross-region routing - create latency records for your resources in multiple regions.
Route53: Failover Routing	Routes to a primary resource when it is healthy or to a secondary when the primary is not healthy.
Route53: Geolocation Routing	Lets you choose the resources that serve your traffic based on the geographic location of your users.
Route53: Geo-proximity Routing	Use the Traffic Flow UI to specify a location for each region, along with a bias
Route53: Multi-value Answer Routing	Returns multiple values (e.g. multiple IP addresses) for your web servers

Route53: Traffic Flow	GUI that lets you build complex routing policies by chaining rules, turning on health checks, etc	
Route53: Alias Record	Route53 alias records are a Route53-specific extension to DNS functionality	
Route53 Resolver	Unifies DNS resolution in a hybrid cloud implementation:	
Route53 Health Checks	Health checks monitor the health and performance of your resources.	
Route 53 pricing	Small monthly cost for each zone	
IAM	Identity and Access Management	
IAM Identities	- Account root user - has full access to everything (and can't be reduced)	
IAM Users	- MFA can be turned on by the user (can't be turned on by the admin for a user)	
IAM Groups	- Policies attached to group apply to all users in the group	
IAM Roles	An IAM identity with specific permissions	
IAM Policies	Managed policies: standalone policy created and administered by AWS; designed to provide permissions for many common use cases (e.g. AmazonDynamo	
IAM Best practices	- Don't create access keys for your AWS account root user unless you absolutely need to	
CloudWatch	Monitoring and observability for cloud applications	
CloudWatch: Dashboards	Create re-usable graphs to visualize your cloud resources in a unified view	
CloudWatch: Events	- can respond to an action (like a change in the AWS environment, or somebody using root credentials to sign in)	
CloudWatch: Alarms	Allow you to set a threshold on metrics and trigger an action	Allow you to set a
CloudWatch: Logs	Monitor, store, and access log files from EC2 instances, CloudTrail, Route53 and other sources	
CloudWatch: Metrics	- collects metrics from more than 70 AWS services (including EC2, DynamoDB, S3, etc) with no action on your part	
CloudWatch: agent	Installed on the guest OS	
CloudWatch pricing	- monthly cost per metric	
CloudTrail	continuous monitoring of activity across your AWS infrastructure	
CloudTrail: Events	Activity is recorded as CloudTrail events:	
CloudTrail: CloudWatch Alarms	CloudWatch alarms can be tied to CloudTrail metrics to alert you when specific actions are taken (e.g. a bucket policy is changed)	
CloudTrail: Security	Logs can be encrypted using KMS	
CloudTrail: Athena	Use Athena to analyze CloudTrail logs	
Kinesis	AWS solution for collecting, processing, analyzing real-time streaming data	
Kinesis Data Streams	Durability: streaming data is replicated across three AZs; data is stored for 7 days	
Kinesis Firehose	Data immediately disappears once processed	
Kinesis Data Analytics	Allows you to perform queries in real-time against a Data Stream or Firehose input	
Kinesis Video Streams	Secuely ingests and stores video and audio , feeding it to consumers such as SageMaker, Rekognition, or other services	
CloudFront	Cloudfront is AWS's CDN, primarily used for speeding up websites by providing cached static content to users at the edge	
CloudFront edge locations	Over 200 points of presence.	
CloudFront web distributions	Specifies things like:	
CloudFront: writing to distributions	A distribution can be enabled to accept POST, PUT, DELETE, OPTIONS, and PATCH requests, allowing you to write content to the distribution	
CloudFront: cache behaviors	A distribution has a default cache behavior, and you can add additional ones (e.g. a cache behavior that applies to images)	
CloudFront: pricing	Customer is charged for the following:	
CloudFront: distribution types	- web distribution (static web content)	
CloudFront: origin access identity	OAI allows CloudFront to serve from non-public S3 buckets	
CloudFront: restricted content	restrict access to files for selected users, for example, users who have paid a fee.	
CloudFront: signed cookies	Signed cookies are best in these cases:	
CloudFront: signed URLs	Signed URLs are best in these cases:	
CloudFront: Lambda@Edge	Allows you to execute functions at the edge to customize the content delivered by CloudFront	
CloudFront: distribution settings	- use an AWS WAF Web ACL	
SQS	SAS is a fully managed message queuing service (like Sidekiq or RabbitMQ)	
SQS: queue types	- standard queues: maximum throughput, best-effort ordering, at-least-once delivery	
SQS: security	- use IAM to grant permission to read from and write to specific queues	
SQS: pricing	Purely based on number of requests	
SQS: long polling	Reduce extraneous polling and receive new messages as quickly as possible.	
SQS: retries, DLQ	If the processing fails, the visibility timeout will expire and the message will be available again	
SQS: visibility timeout	When a consumer receives and processes a message from a queue, the message remains in the queue	
SQS: message retention	SQS can retain messages from 60 seconds to 14 days	
SQS vs Amazon MQ	Amazon MQ is a managed Apache ActiveMQ service and supports queue and broadcast protocols like AMQP, JMS, etc.	
SQS vs SNS	SNS is a distributed pub-sub system. All messages are pushed to all receivers	
SNS	Fully managed pub/sub messaging service	
SNS: supported receivers	- HTTP/HTTPS endpoints	
SNS: security	- use IAM to grant permission to subscribe to and publish to specific topics	
SNS: retries, DLQ	If an endpoint is not available, SNS will execute a retry policy.	
SNS: availability	Data is replicated across AZs	
SNS: AWS services with publishing su	Over 30 services support publishing to SNS, including:	
SNS: message filtering	Subscribers can specify a filter policy so that it only gets a subset of the messages posted to a topic	
SNS: pricing	Strictly pay-as-you go	
CloudFormation	CloudFormation is a templating system which allows you to automate the provisioning of AWS resources and third-party resources	
CloudFormation: devops	Can house your CloudFormation templates in a git repository and deploy to your AWS environment via a CI/CD pipeline	
CloudFormation: NestedStacks	Allow you to break up your template into smaller reusable templates that can be composed together into a larger template	
CloudFormation: template anatomy	- Description: text string describing the template	
SWF	Simple Workflow Service	

SWF functionality	- maintains application state
SWF implementation	- define a workflow
OpsWorks	Configuration management service that provides managed instances of Chef and Puppet
Well-Architected Framework	Pillars:
Well-Architected Framework:	- perform operations as code
Well-Architected Framework:	- implement a strong identity foundation
Well-Architected Framework:	- automatically recover from failure
Well-Architected Framework:	- democratize advanced technologies
Well-Architected Framework:	- implement cloud financial management
Well-Architected Framework:	- stop guessing your capacity needs
VPC Endpoints	Allow you to communicate with supported AWS services without the traffic traversing the public network (it all stays inside your VPC)
VPC Endpoints: Interface endpoint	An interface VPC endpoint (interface endpoint) enables you to connect to services powered by AWS PrivateLink.
VPC Endpoints: Gateway endpoint	A gateway that you specify as a target for a route in your route table for traffic destined to a supported AWS service.
PrivateLink	Provides private connectivity between VPCs and services hosted on AWS or on-premises, securely on the Amazon network.
EFA	An EFA is an Elastic Network Adapter (ENA) with added capabilities. It provides all of the functionality of an ENA, with an additional OS-bypass functionality.
Elastic Beanstalk	An orchestration service for applications leveraging various AWS services, including EC2, S3, SNS, CloudWatch, autoscaling, and Elastic Load Balancers
Elastic Beanstalk: supported language	Go, Java, .NET, Node.js, PHP, Python, and Ruby
Elastic Beanstalk: components	- application: a logical container for the project
Security Token Service (STS)	A web service that enables you to request temporary, limited-privilege credentials for IAM users or for users that you authenticate (federated users)
DataSync: Verification	AWS DataSync locally calculates the checksum of every file in the source file system and the destination and compares them.
Server Migration Service	Automates the migration of your on-premises VMware vSphere, Microsoft Hyper-V/SCVMM, and Azure virtual machines to the AWS Cloud.
Autoscaling: termination policies	Governs the selection of instances to terminate in a scale-in situation
Autoscaling: lifecycle hooks	Lifecycle hooks enable you to perform custom actions by pausing instances as an Auto Scaling group launches or terminates them.
Autoscaling: mixing purchase options	An ASG can be comprised of on-demand and spot instances.
Organizations	Organizations helps you to centrally manage billing; control access, compliance, and security; and share resources across your AWS accounts.
Organizations: Resource Access Man	RAM lets you share your resources with any account or through Organizations.
Organizations: Service Control Policie	SCPs are a type of organization policy that you can use to manage permissions in your organization
CloudFormation: drift detection	Detects any changes to the current stack configuration to the one specified in the template and reports differences.
EC2: billing (hours/seconds)	Per-second billing is available for Linux instances; per-hour for all other instance types
AWS Global Accelerator	Global Accelerator uses the AWS global network to route traffic to the optimal regional endpoint based on health, client location, and policies that you configure.
AWS Global Accelerator vs Route 53	With Route 53 geolocation, you have to deal with DNS caching; Global Accelerator doesn't use DNS to route traffic.
Redshift	Amazon Redshift is a fully-managed petabyte-scale cloud based data warehouse product designed for large scale data set storage and analysis. It is also used to perform large scale database migrations.
AWS Key Management Service	KMS is a secure and resilient service that uses hardware security modules that have been validated under FIPS 140-2 (or are in the process of being validated) to protect your keys.
S3: Storage Class Analysis	Lets you analyze storage access patterns to decide when to transition the right data to the right storage class.
SQS: delay queue	Delay queues let you postpone the delivery of new messages to a queue for a number of seconds, for example, when your consumer application needs additional time to process messages. If you create a delay queue, any messages that you send to the queue remain invisible to consumers for the duration of the delay period.
Elastic Container Service	ECS is a fully managed container orchestration service.
Elastic Container Registry	A fully-managed Docker container registry, integrated with ECS
Fargate	Fargate is a serverless compute engine for containers that works with both Amazon Elastic Container Service (ECS) and Amazon Elastic Kubernetes Service (EKS).
CloudTrail: cross-region	You can turn on a trail across all regions -- CloudTrail will deliver log files from all regions to the S3 bucket and an optional CloudWatch Logs log group you specified.
FSx	Fully managed filesystems
FSx for Windows File Server	Provides fully managed file storage that is accessible over the industry-standard Server Message Block (SMB) protocol.
FSx for Lustre	Lustre is a type of parallel distributed file system, generally used for large-scale cluster computing.
VPC Flow Logs	A feature that enables you to capture information about the IP traffic going to and from network interfaces in your VPC.
VPC Peering	A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses.
AWS Config	AWS Config is a service that enables you to assess, audit, and evaluate the configurations of your AWS resources.
Cognito	Cognito lets you add user sign-up, sign-in, and access control to your web and mobile apps quickly and easily.
Cognito: User Pools	A user pool is a user directory in Cognito. With a user pool, your users can sign in to your web or mobile app through Cognito.
Cognito: Identity Pools	Identity pools provide temporary AWS credentials for users who are guests (unauthenticated) and for users who have been authenticated and received a token.
Cognito: social identity providers	Users can sign in through social identity providers (IdP) like Facebook, Google, Amazon, and Apple.
S3: Presigned URLs	Allows the owner to share private objects with others by creating a presigned UR to grant time-limited permission to download the objects.
ClassicLink	ClassicLink allows you to link an EC2-Classic instance to a VPC in your account, within the same region. This allows you to associate the VPC security groups with the EC2-Classic instance, enabling communication between your EC2-Classic instance and instances in your VPC using private IPv4 addresses.
Database Migration Service	a cloud service that makes it easy to migrate relational databases, data warehouses, NoSQL databases, and other types of data stores.
Database Migration Service: Schema	The Schema Conversion Tool makes heterogeneous database migrations predictable by automatically converting the source database schema and a majority of the database code objects, including views, stored procedures, and functions, to a format compatible with the target database.
Glacier Select	Allows you to perform filtering operations using simple Structured Query Language (SQL) statements directly on your data in S3 Glacier.
S3 Select	Filters the contents of an S3 object based on a simple structured query language (SQL) statement.
Glacier: restore	Restoring from Glacier and Glacier Deep Archive creates a temporary copy of the object in S3 (duration is user-specified).
Serverless Application Model	SAM is an open-source framework for building serverless applications.
EC2: Hibernation	Hibernation saves the contents from the instance RAM to the EBS root volume. EC2 persists the instance's EBS root volume and any attached EBS data volumes.
Secrets Manager	Service that encrypts and stores your secrets, and transparently decrypts and returns them to you in plaintext.
AWS Budgets	AWS Budgets allows you to set custom budgets to track your cost and usage from the simplest to the most complex use cases.
Redshift: Spectrum	Redshift Spectrum is a feature within Redshift that lets a data analyst conduct fast, complex analysis on objects stored in S3.
Inspector	An automated security assessment service that helps improve the security and compliance of applications deployed on AWS.
Aurora: endpoints	- cluster endpoint (aka "writer endpoint"): connects to the current primary DB instance for that DB cluster
Macie	A fully managed data security and data privacy service that uses machine learning and pattern matching to discover and protect your sensitive data in AWS.
RDS: enhanced monitoring	Amazon RDS provides metrics in real time for the operating system (OS) that your DB instance runs on, using an agent installed on the instances.
Autoscaling: default policy	- filters for instances in the AZ with the most instances.

ECS: task definitions	A task definition is required to run Docker containers in ECS. Some of the parameters you can specify in a task definition include:										
Health Checks: ELB vs ASG	Auto Scaling can use these health checks:										