AWS CSA points to remember

Limits:

- Default 20 EC2 per region
- Default 5000 volumes or 20 TiB
- 5 EIP per region
- 10 ELBs
- Default 2 high I/O instances
- Default 5 VPCs per region (5 IGWs per region as 1 IGW can attach to 1 VPC)
- 50 customer gateways & 50 VPN connections per region
- 200 RTs per region; 50 entries per RT
- 5 SG per ENI, 50 rules per SG
- EC2 classic : Max 500 SGs per region and SG can have max 100 rules
- EC2-VPC : Up to 100 SGs per VPC

S3:

- 0 bytes to 5 TB files
- Multipart upload is recommended for objects larger than 100MB BUT required for greater than 5GB
- S3 & S3-IA → 99.999999999% durability

  S3-IA → 99.9% availability

  S3 → 99.99% availability
- 100 buckets/account. Bucket ownership cannot be transferred once bucket is created
- Types of S3 → S3, S3-IA (infrequent access), Glacier, RRS
- RRS → 99.99% durability
- S3 versioning can be suspended but cannot be disabled.
- ACL is used to share bucket across accounts & bucket policies is used to share bucket with anonymous users
- Cloudfront + S3 = Signed URL with time limit access.
- S3 encryption - SSE, SSE-C, SSE-KMS, Encryption client library.

VPC:

- Classless Internet Domain Routing (CIDR) block acts as router. Default VPC has CIDR 172.31.0.0/16
- VPC size → /28 to /16 in IPv4 & /56 in IPv6
- A subnet has 251 available IPs by default
- RT → Dest : 0/0, Target : IGW id for public subnet
- SG - Works at VPC/instance layer (use to block certain ports)
  Stateful in nature
- Network ACL - Works at network/subnet level (use to block entire network)
  Stateless in nature
- VPC peering can be done between 2 VPCs in the same region
- Invalid VPC peering connections → Overlapping CIDR blocks, Transitive peering,
  Edge to Edge Routing Through a Gateway or Private Connection (VPC A<-->VPC B, VPC A<-->VPN to corporate NW, Corporate NW <-/-> VPC B)
- IGW is horizontally-scaled, redundant, HA with no bandwidth constraints.

EC2:

- T2 instances are burstable performance instances which accumulate CPU credits when idle & use them when active. A stopped instance does not retain its previously earned credit balance.
- 1 CPU credit = 100% CPU usage for 1 minute.
- T2.micro (1 vCPU) → default 6 CPU credits per hour
  → max 144 CPU credits
  → can use 10% of core at base CPU performance
- EBS types → SSD, PIOPS, Magnetic
- EBS snapshot is incremental in nature and can be reverted back to point in time
- Enhanced Networking is only supported in VPC
- EBS snapshot can be done in real time but new data being written on volume will NOT be included in snapshot. Same time is taken to snapshot a volume of 16TB and 1TB
- PIOPS volumes are billed even if they are detached, it's best to take their snapshot and delete the original volume
- Classic ELB → simple LB of traffic is to be done across multiple EC2s
  Application LB → Use when app needs advanced routing capabilities for
  microservices & container based architectures
- Variable attributes while reserving an instance are instance type, OS, tenancy, payment option (region cannot be changed)
- If a spot instance is terminated by AWS before an hour → no charge for partial usage
  BUT if spot instance is terminated by us before an hour → charged for 1 hour

IAM:

- User ARN → arn:aws:iam::<AWS account no.>:user/<username>
- IAM + Cloudtrail → Used to maintain user logs on AWS by tracking API key of user
- IAM users can have creds like AWS access keys, X.509 cert, ssh keys, password for login & MFA devices.
- IAM roles are meant to be assumed by authorized entities (users/apps/services) within or between AWS accounts
- By default 250 IAM roles can be made

RDS:

- Multi AZ is a high availability feature and NOT a scaling solution. For read-only scenarios, standby replica cannot be used to serve read traffic. To service read-only traffic use READ REPLICAS.
- Failover mechanism automatically changes the DNS record of the DB instance to point to the standard DB instance.
- Benefits of running RDS instead of EC2 with DB :
    a. Automatic minor updates
    b. Automatic backups
    c. Not required to manage OS
    d. Multi AZ with single click
    e. Automatic recovery in event of failover
- MySQL requires InnoDB for backups
- Automated backups
    a. RDS uses periodic data backups + transaction logs to enable restoration of DB instance to any second during the restoration period, upto the LatestRestorableTime.
    b. No I/O suspension for multi AZ instance (single AZ experiences it)
- In manual backups, RDS keeps all backups until deleted explicitly
- Failover mechanism auto changes the DNS record of DB instance to point to standard instance.
- MySQL databases have only one license model i.e. general-public-license. Other proprietary databases such as Oracle & SQL servers offer two license modes - Licenses Included & BYOL

Route 53:

- Route 53 + CloudFormation = Setup duplicate architecture in JSON template
- Use routing policies to manipulate load on domain. Routing policies are - Simple, Weighted, Latency, Failover, Geological.
- Create A record alias for ELB to host a domain in route 53
- Supported DNS resource record types →

    - A - IPv4 address
    - AAAA - IPv6 address
    - CNAME - domain name format. Cannot be created for zone apex but can be made for subdomain (ex - failover.example.com)
    - MX - mail xchange format

CloudFront:

- It's Content Delivery Service (CDN)
- The origin can be S3 endpoint or ELB's CNAME. The content is cached at edge location. Create Origin Access Identity (OAI) to restrict access of content hence preventing users to directly accessing the bucket.

SQS:

- SQS is message oriented API
- It helps build distributed application with decoupled components
- 1 message = 256KB of text in any format
- SQS asynchronously PULLS the task messages from queue (polling)
- No guarantee if FIFO for messages, just that a message will get delivered at least once. Makes best effort to preserve order of messages.
- SQS stores the message up to 4 days, by default, and can be configured from 1 minute to 14 days but clears the message once deleted by the consumer
- The maximum visibility timeout for an Amazon SQS message is 12 hours
- To avoid polling in tight loops (which burns CPU cycles) enable long polling by setting the ReceiveMessageWaitTimeSeconds to > 0 which reduces number of empty responses

SWF:

- SWF is task oriented API
- It coordinates work across distributed application components
- Ensures task is assigned only once and never duplicated
- A workflow execution can last up to 1 year

SNS:

- SNS + CloudWatch = Environment monitoring
- Create topic → create subscription → email address to send notifications on as endpoint
- Subscription can be HTTP/HTTPS/SMS/EMAIL/EMAIL-JSON/SQS or application
- Uses PUSH based mechanism (unlike SQS which uses pull/polling mechanism)

Kinesis:

- Stores records of a stream for up to 24 hours by default, which can be extended to maximum 7 days
- Can export data to services like EMR, S3, Redshift, Lambda
- Benefits - real-time processing, parallel processing, durable, scalability
- Use cases - gaming, real-time analytics (IOT), application alerts, log/event data collection, mobile data capture

- Producers create data (sensors, mobile devices) & Consumers consume data (S3, EMR, Redshift, Lambda)
- Amazon Kinesis Producer Library (KPL) is an easy to use and highly configurable library that helps you put data into an Amazon Kinesis stream
- Amazon Kinesis Client Library (KCL) is a pre built library which acts as an intermediary between your record processing logic and Streams

EMR:

- Gives admin ability to access underlying OS
- Integrates with S3, DynamoDB, RedShift to send and receive data
- Mapper looks out for occurrence of sample data whereas Reducer takes the output and from mapper & brings it down to single single output file

Elastic Beanstalk:

- Supported platforms → Docker, Java, Windows .NET, Node.js, PHP, Python, Ruby
- Little to no management required to deploy single-tier applications
- Whatever templates we build in elastic beanstalk are displayed in CloudFormation

CloudFormation:

- Infrastructure as Code
- Templates are built in JSON format

Miscellaneous:

- ➔ As a part of shared responsibility model, AWS provides security for Physical network infrastructure, Virtualization Infrastructure & Regulatory infrastructure
- ➔ /28 is the smallest possible subnet in an AWS VPC
- ➔ SQS & SWF can be used as an API for on-premises hardware too
- ➔ While designing an arch with EC2 & ELB to determine instance size we must know the minimum memory requirements for the application & required I/O operations
- ➔ Decommissioning of storage devices using industry standard practices is an operational process performed by AWS for data security
- ➔ KPL acts as an intermediary between producer application code & stream's API actions. Aggregation in KPL refers to the storage of multiple records in a stream's record and allows customers to increase the number of records sent per API call, which effectively increases producer throughput.

Q. Why I route 53 named so?
Q. Which service integrates with Chef right out of the box (OpsWorks)
Q. User desktop availability over the cloud (AWS WorkSpaces)
Q. Multiple regions with multiple AZs, how will you design HA & scalable infra if each each instance can handle 45% CPU load? (scenario based)
Q. Where do you change source/destination check (NAT instance)
Q. What option is to be enabled to equally distribute load across AZs below an ELB (cross-zone load balancing)