“100 AWS Interview Questions & Answers – Crack Cloud Job Interviews with Confidence”

✉️ A Note from the Author

Dear Reader,

Thank you for downloading this guide!

As someone deeply involved in cloud and data engineering, I know firsthand how overwhelming AWS interview preparation can feel — with its vast set of services, real-world use cases, and syntax-heavy solutions. That’s why I created this product: to help you bridge the gap between theory and practice.

In this guide, I’ve included not only interview questions and answers, but also:

- Real-world scenarios you might actually face on the job
- AWS CLI and service-specific syntax and examples
- Step-by-step architecture walkthroughs

Whether you're a fresher or an experienced engineer aiming for your next big cloud role — my goal is to help you understand, not just memorize.

I hope this becomes a valuable asset in your journey to crack AWS interviews with confidence and grow as a cloud professional.

Wishing you all the best 🚀

– Suden Gorai
Author & Cloud Enthusiast

✅ **Structure:**

🔹 Part 1: Basic AWS Concepts

✅ Q1: What is AWS?

Answer:
Amazon Web Services (AWS) is a **cloud computing platform** provided by Amazon that offers on-demand services like **compute power, storage, networking, databases**, and more — on a pay-as-you-go model.

Example:
Instead of buying your own server, you can rent a virtual one (EC2) on AWS for only the hours you need it.

✅ Q2: What are the key benefits of using AWS?

Answer:

- **Scalability:** Automatically adjust resources to match demand
- **Pay-as-you-go:** No upfront costs
- **High availability:** Global data centers (Availability Zones)
- **Security:** Data encryption, IAM, compliance standards
- **Global reach:** 30+ regions across the world

✅ Q3: What is the difference between IaaS, PaaS, and SaaS?

Answer:

- **IaaS (Infrastructure as a Service):** You manage software; AWS provides infrastructure (e.g., EC2)
- **PaaS (Platform as a Service):** AWS manages infrastructure & runtime (e.g., AWS Elastic Beanstalk)
- **SaaS (Software as a Service):** Fully managed applications (e.g., Amazon Chime)

✅ Q4: What is the AWS Free Tier?

Answer:
The AWS Free Tier allows new users to explore AWS for **12 months** with limited free access to services like EC2, S3, RDS, Lambda.

Example:
You can use a **t2.micro EC2 instance** with 750 hours/month for free in the first year.

✅ Q5: What are Availability Zones and Regions?

Answer:

- A **Region** is a geographical area (e.g., us-east-1)
- An **Availability Zone (AZ)** is a data center within a region
- AZs offer redundancy and fault tolerance

Example:
us-east-1 has multiple AZs (like us-east-1a, 1b, 1c) — you can deploy your app in multiple AZs to ensure high availability.

✅ Q6: What is Amazon S3?

Answer:
Amazon S3 (Simple Storage Service) is **object storage** used for storing unstructured data (files, images, videos, etc.).

Example:
You can host a static website or store user profile images in S3.

✅ Q7: What is EC2?

Answer:
Amazon EC2 (Elastic Compute Cloud) allows you to rent virtual servers (instances) to run applications.

Example:
Use EC2 to deploy a backend service like a Django or Node.js API.

✅ Q8: What is IAM?

Answer:
IAM (Identity and Access Management) is a service to **securely manage access** to AWS services and resources.

Example:
You can create a user “developer1” who only has read access to S3 but no access to EC2.

✅ Q9: What is a VPC?

Answer:
A VPC (Virtual Private Cloud) is your own **isolated network** in AWS where you can launch EC2 instances, create subnets, and control traffic flow.

Example:
You can create a public subnet for a web server and a private subnet for a database server within the same VPC.

✅ Q10: What is an Elastic IP?

Answer:
An Elastic IP is a **static, public IPv4 address** that you can allocate to your AWS account and associate with EC2 instances.

Example:
If you want your EC2 server to have a **permanent public IP** (instead of dynamic), use an Elastic IP.

🔷 Part 2: Core AWS Services (Questions 11–30)

✅ Q11: What is the difference between EC2 and Lambda?

Answer:

- **EC2**: You manage the server. Good for long-running, stateful applications.
- **Lambda**: Serverless — AWS runs the code for you in response to events. Ideal for short, stateless executions.

Example:
Use EC2 to host a backend API. Use Lambda to process image uploads to S3.

✅ Q12: How do EC2 instance types differ?

Answer:
EC2 instances are categorized by use case:

- **t2/t3**: General purpose (web servers, dev)
- **m5**: Balanced compute and memory
- **c5**: Compute-optimized (batch processing)
- **r5**: Memory-optimized (databases)
- **g4**: GPU-based (machine learning)

✅ Q13: What is EBS in AWS?

Answer:
EBS (Elastic Block Store) is **block storage** for EC2. Like a virtual hard disk.

Example:
When you launch an EC2 instance, its root volume is typically an EBS volume.

✅ Q14: What are the different storage classes in S3?

Answer:

- **Standard**: Frequent access
- **IA (Infrequent Access)**: Less frequent, lower cost
- **Glacier**: Archival, retrieval time in minutes/hours
- **One Zone-IA**: Like IA but stored in one AZ
- **Intelligent-Tiering**: Auto-moves between tiers

✅ Q15: How is data secured in S3?

Answer:

- **Encryption**: SSE-S3, SSE-KMS, or client-side
- **Access Control**: IAM policies, Bucket policies, ACLs
- **Versioning**: Protects against accidental deletion

✅ Q16: What is an S3 Bucket Policy?

Answer:
A **JSON-based access policy** attached to an S3 bucket to control permissions.

Example:
Allow all public users to read objects in a static website bucket.

✅ Q17: What is Amazon RDS?

Answer:
RDS (Relational Database Service) is a managed database service for SQL-based databases like MySQL, PostgreSQL, Oracle, SQL Server, and Aurora.

Example:
You can deploy a MySQL database without managing backups, patching, or high availability.

✅ Q18: What is Multi-AZ in RDS?

Answer:
Multi-AZ provides **automatic failover** for high availability. AWS maintains a standby replica in a different AZ.

✅ Q19: What is Read Replica in RDS?

Answer:
Read Replicas allow you to **scale read operations** by duplicating your database in real-time.

Example:
Use replicas for analytics queries while keeping the master DB fast for writes.

✅ Q20: What is Amazon Aurora?

Answer:
Aurora is a high-performance, fully managed MySQL/PostgreSQL-compatible database with better performance and scalability than traditional RDS.

✅ Q21: What is Amazon Lambda?

Answer:
Lambda is a **serverless compute service** that runs code in response to triggers (like S3 uploads, API Gateway, or CloudWatch events).

Example:
Automatically resize images uploaded to an S3 bucket.

✅ Q22: How does Lambda handle scaling?

Answer:
Lambda **auto-scales** based on the number of incoming requests. Each request triggers a separate execution environment.

✅ Q23: What is the max execution time of a Lambda function?

Answer:
Currently, **15 minutes** is the maximum allowed runtime per execution.

✅ Q24: What is API Gateway?

Answer:
Amazon API Gateway allows you to **build and expose REST APIs or WebSocket APIs** and connect them to Lambda, EC2, or other services.

Example:
Use API Gateway + Lambda to build a serverless backend.

✅ Q25: What is AWS CloudWatch?

Answer:
CloudWatch is a monitoring service to track **metrics, logs, and alarms** for AWS resources.

Example:
Set up an alarm if EC2 CPU utilization goes above 80%.

✅ Q26: What is AWS CloudTrail?

Answer:
CloudTrail records **API calls** and activity across AWS services for **audit and compliance**.

Example:
Track who deleted an S3 bucket or launched a new EC2 instance.

✅ Q27: What is an Auto Scaling Group (ASG)?

Answer:
ASG automatically **scales EC2 instances** up/down based on demand (CPU usage, etc.).

✅ Q28: What is Elastic Load Balancer (ELB)?

Answer:
ELB automatically distributes incoming traffic across multiple EC2 instances.

Types:

- Application Load Balancer (HTTP/HTTPS)
- Network Load Balancer (TCP)
- Gateway Load Balancer (3rd-party appliances)

✅ Q29: What is Route 53?

Answer:
Route 53 is AWS's **DNS and domain management service**. Supports routing, health checks, failover, etc.

✅ Q30: What is AWS CloudFormation?

Answer:
CloudFormation allows you to define your infrastructure using **templates (YAML/JSON)**, enabling "Infrastructure as Code".

Example:
Automate deployment of a VPC + EC2 + RDS with one file.

🔷 Part 3: Scenario-Based & Advanced AWS Interview Questions (Q31–60)

These questions test **hands-on experience, decision-making**, and how well you understand real AWS architectures.

✅ Q31: How would you make a highly available web application on AWS?

Answer:
Use a combination of services:

- Deploy the app on **EC2 Auto Scaling Groups** across **multiple AZs**
- Use an **Application Load Balancer (ALB)** to route traffic
- Store static files in **S3**
- Use **RDS with Multi-AZ** or **Aurora**
- Place app servers in **private subnets**, public ALB in public subnet
- Use **Route 53** for DNS + health checks

✅ Q32: Your EC2 instance is unreachable. What steps do you take to troubleshoot?

Answer:

1. Check **Instance State** and **System Status Checks** in EC2
2. Verify **Security Group** (port 22 or 80 open)
3. Check **Network ACLs**
4. Ensure EC2 is in **public subnet with Internet Gateway**
5. Check **Elastic IP** or public IP
6. Review **CloudWatch logs** or EC2 logs if enabled

✅ Q33: What happens if you delete the root EBS volume of an EC2 instance?

Answer:
If the **Delete on Termination** flag is set to true, the root volume will be deleted when the instance is terminated. If set to false, it will persist after instance termination.

✅ Q34: You want to store 100TB of infrequently accessed data. Which S3 storage class should you choose?

Answer:
Use **S3 Glacier Deep Archive** or **S3 Standard-IA** depending on retrieval needs. Glacier is cheaper but slower for access.

✅ **Q35: How can you automate daily backups for RDS?**

Answer:

- Enable **automated backups** (daily snapshot + transaction logs)
- Set **retention period** up to 35 days
- Alternatively, create custom **scheduled Lambda + snapshot** logic for more control

---

✅ **Q36: How do you restrict an IAM user to access only a specific S3 bucket?**

Answer:
Create an **IAM policy** like:

```
{
 "Effect": "Allow",
 "Action": "s3:*",
 "Resource": [
  "arn:aws:s3:::your-bucket",
  "arn:aws:s3:::your-bucket/*"
 ]
}
```

Then attach it to the IAM user.

✅ **Q37: How can you transfer 1 TB of data from on-prem to AWS securely?**

Answer:

- Use **AWS Snowball** (for large scale, offline transfer)
- Or use **AWS DataSync** or **S3 Transfer Acceleration**
- Use **KMS** or client-side encryption for sensitive data

✅ **Q38: You need to serve a static website with low latency worldwide. What services do you use?**

Answer:

- **Amazon S3** to host the site
- **CloudFront** (CDN) to cache content close to global users

- **Route 53** to map domain to CloudFront distribution

✅ **Q39: How do you make Lambda functions access resources in a private VPC?**

Answer:

- Attach the **Lambda function to the VPC**
- Choose private subnets with a **NAT Gateway** if it needs outbound internet
- Set up proper **Security Groups**

✅ **Q40: What is the best way to control cost across multiple AWS accounts in an organization?**

Answer:

- Use **AWS Organizations** to manage billing
- Set **budgets and alerts** using **AWS Budgets**
- Use **Cost Explorer** for visualizations
- Use **Service Control Policies (SCPs)** to restrict expensive services

---

✅ **Q41: How does AWS handle eventual consistency in S3?**

Answer:

- **Read-after-write consistency** for new PUTs to new objects
- **Eventual consistency** for overwrite PUTs and DELETEs
- Meaning: reading immediately after a delete might still return the old object briefly

---

✅ **Q42: What is a NAT Gateway?**

Answer:
A **NAT Gateway** allows instances in a **private subnet** to connect to the internet **without receiving inbound traffic** from the internet.

✅ **Q43: What is the difference between NACL and Security Groups?**

Answer:

Feature	NACL	Security Group
Type	Stateless	Stateful

Feature	NACL	Security Group
Rules Applied To	Subnets	EC2 Instances
Allow/Deny Support	Allow and Deny	Allow only
Rule Evaluation	By rule number (lowest wins)	All rules applied

---

✅ **Q44: How do you enable high availability for a web server?**

Answer:

- **ALB + Auto Scaling Group** across multiple AZs
- Store session data in **ElastiCache or DynamoDB**
- Use **Route 53 with failover routing**
- Use **multi-region disaster recovery** if needed

✅ **Q45: What is a VPC peering connection?**

Answer:
VPC Peering allows **network traffic** between two VPCs using **private IPs**.

Example:
Connect VPC-A and VPC-B without needing a VPN or NAT Gateway.

✅ **Q46: What is AWS KMS?**

Answer:
AWS KMS (Key Management Service) is used to create and manage **encryption keys** for your AWS resources.

✅ **Q47: Can Lambda functions run in multiple AZs?**

Answer:
Yes. Lambda is a **fully managed service** that runs code in multiple AZs automatically. No need to choose one.

✅ **Q48: You need to run a script every day at 8 AM. How?**

Answer:
Use **EventBridge (CloudWatch Events)** to trigger a **Lambda** function on a **scheduled cron expression**.

✅ **Q49: How can you share a snapshot of your RDS database with another AWS account?**

Answer:
Make the snapshot **public or share it** with a specific AWS account under snapshot permissions.

✅ **Q50: How can you protect against accidental deletion of S3 objects?**

Answer:

- Enable **versioning**
- Use **MFA Delete**
- Set **bucket policies** to deny delete actions unless specific conditions are met

🔷 Part 4: Security, DevOps, Billing & Monitoring (Q51–80)

These questions focus on topics commonly asked in **DevOps**, **Cloud Security**, and **Cloud Cost Management** roles.

✅ Q51: What is the Shared Responsibility Model in AWS?

Answer:

- **AWS** is responsible for security **of** the cloud (infrastructure, hardware, data centers).
- **You (the customer)** are responsible for security **in** the cloud (IAM, S3 access, encryption, etc.).

✅ Q52: How does AWS KMS help secure your data?

Answer:
AWS Key Management Service (KMS) enables you to create and manage encryption keys for S3, RDS, Lambda, etc.

Example:
Use KMS to encrypt S3 objects with a customer-managed key.

✅ Q53: What is Secrets Manager?

Answer:
AWS Secrets Manager stores and rotates **secrets** like database passwords, API keys, and tokens securely.

Bonus: It supports automatic rotation using Lambda.

✅ Q54: How is CloudTrail different from CloudWatch?

Answer:

- **CloudTrail** logs all API calls (who did what, when) – ideal for auditing.
- **CloudWatch** monitors metrics (CPU, memory, logs) – ideal for system health.

✅ Q55: How do you protect an S3 bucket from public access?

Answer:

- Block public access at bucket level
- Remove any bucket policies or ACLs that allow public access
- Use IAM policies to restrict access

✅ Q56: What is an IAM Role and how is it different from a User?

Answer:

- An **IAM User** has long-term credentials (username + password / access keys)
- An **IAM Role** is assumed by AWS services or federated users and has **temporary security credentials**

Example: EC2 assuming a role to access S3:

```
{
 "Effect": "Allow",
 "Action": "s3:ListBucket",
 "Resource": "arn:aws:s3:::my-bucket"
}
```

Attach this IAM Role to the EC2 instance, and it can access S3 **without embedding credentials**.

✅ Q57: What is AWS Config?

Answer:
AWS Config **monitors and records changes** to AWS resources for audit/compliance.

Example CLI to enable config recorder:

```
aws configservice start-configuration-recorder \
 --configuration-recorder-name default
```

✅ Q58: What are Service Control Policies (SCP) in AWS Organizations?

Answer:
SCPs allow central control of **permissions across accounts**. You can restrict access to services like EC2 or S3 across an entire org unit.

✅ Q59: How can you monitor the cost of your AWS usage?

Answer:

- Use **AWS Budgets** to set spending limits and get alerts
- Use **Cost Explorer** for visual cost analysis
- Use **Billing Dashboard** for detailed breakdowns

✅ Q60: What is Consolidated Billing?

Answer:
It allows **multiple AWS accounts** in an AWS Organization to share one billing account for **bulk discount pricing** and simplified management.

✅ Q61: What is Amazon Inspector?

Answer:
It is a security assessment service that automatically scans EC2 instances or containers for **vulnerabilities and compliance issues**.

✅ Q62: What is GuardDuty?

Answer:
A threat detection service that uses machine learning to identify suspicious behavior like **unauthorized access or data exfiltration**.

✅ Q63: What is AWS WAF?

Answer: WAF protects web apps by filtering malicious traffic.
Example rule to block SQL injection:

```
{
 "Name": "BlockSQLInjection",
 "Priority": 1,
 "Action": {"Block": {}},
 "Statement": {
  "SqliMatchStatement": {
   "FieldToMatch": {"Body": {}},
   "TextTransformations": [{"Priority": 0, "Type": "URL_DECODE"}]
  }
 }
}
```

Deploy this rule via **WAF WebACL** and associate it with an ALB or CloudFront distribution.

✅ Q64: What is Shield and how does it differ from WAF?

Answer:

- **Shield**: DDoS protection (Standard is free, Advanced is paid)
- **WAF**: Application-level firewall for web exploits
- They work together to protect apps

✅ Q65: How do you enable CI/CD in AWS?

Answer:
Use the **AWS Developer Tools suite**:

- CodeCommit → Git repo
- CodeBuild → Build server
- CodeDeploy → Deployment manager
- CodePipeline → Orchestration tool for CI/CD

✅ Q66: How can Lambda be used in DevOps pipelines?

Answer:

- Post-deployment validation
- Rollback automation
- Notifying Slack or email after CodePipeline execution

✅ Q67: What is CloudFormation Drift Detection?

Answer:
It checks whether your deployed resources differ from your **CloudFormation template** — helps identify manual changes.

✅ Q68: What is Elastic Beanstalk?

Answer:
A Platform-as-a-Service (PaaS) that handles provisioning, load balancing, and auto-scaling for apps in Node.js, Python, Java, etc.

✅ Q69: What are CloudWatch Logs and Log Insights?

Answer:

- **CloudWatch Logs** store log data from EC2, Lambda, ECS, etc.
- **Log Insights** lets you **query logs** with SQL-like syntax for debugging and monitoring.

Sample Log Insights query (Lambda errors):

```
fields @timestamp, @message
| filter @message like /ERROR/
| sort @timestamp desc
| limit 20
```

✅ Q70: How do you monitor RDS performance?

Answer:

- Use **CloudWatch metrics** for CPU, memory, connections
- Use **Enhanced Monitoring** for OS-level metrics
- Use **Performance Insights** to find slow queries

✅ Q71: What is Spot Instance and when should you use it?

Answer:
Spot Instances are **unused EC2 capacity** at up to 90% discount, but can be **terminated anytime**.

Best for: Batch jobs, big data, non-critical workloads

✅ Q72: What is Elastic Container Service (ECS)?

Answer:
ECS is a **container orchestration service** to deploy and scale Docker containers. Can run on EC2 or Fargate.

✅ Q73: What is Fargate?

Answer:
A **serverless compute engine** for containers. You don't manage servers, just define the container specs.

✅ Q74: How do you enable logging for S3 access?

Answer:

- Enable **server access logging** or **CloudTrail data events**
- Logs are written to another S3 bucket for auditing

Via S3 console or CLI:

```
aws s3api put-bucket-logging \
 --bucket my-bucket \
 --bucket-logging-status '{
   "LoggingEnabled": {
     "TargetBucket": "my-logging-bucket",
     "TargetPrefix": "logs/"
   }
 }'
```

This logs all S3 access events into the my-logging-bucket.

✅ Q75: What is Trusted Advisor?

Answer:
A service that provides **real-time recommendations** on cost optimization, security, performance, and service limits.

✅ Q76: What is Amazon ECR?

Answer:
Amazon Elastic Container Registry is a managed **Docker container registry** for storing, managing, and deploying images.

✅ Q77: What is an Amazon AMI?

Answer:
Amazon Machine Image is a template used to launch EC2 instances. It contains OS, application server, and apps.

✅ Q78: What is EventBridge?

Answer:
A serverless event bus that lets AWS services, SaaS apps, and your code **react to events** across your system.

✅ Q79: What is Step Functions?

Answer:
Step Functions is a serverless workflow service that **orchestrates AWS services** using a visual interface and state machine logic.

✅ Q80: How do you restrict access to Lambda only from a specific VPC?

Answer:

- Use **VPC Endpoint** + **Resource Policy** to restrict Lambda invocations
- Or use **security groups + private subnets**

🔷 Part 5: Multiple-Choice & Situational AWS Interview Questions (Q81–100)

These questions simulate real interview patterns, with **detailed explanations and syntax** to help reinforce concepts.

✅ Q81: Which of the following services is serverless?

A. EC2
B. Lambda
C. RDS
D. Elastic Beanstalk

Correct Answer: B. Lambda

Explanation:
AWS Lambda runs code without provisioning servers. Others require managing servers or containers.

✅ Q82: How can you encrypt data at rest in S3?

A. Use SSL
B. Use IAM Policies
C. Enable Server-Side Encryption (SSE)
D. Use AWS WAF

Correct Answer: C. Enable Server-Side Encryption (SSE)

Example (with SSE-S3):

```
aws s3 cp file.txt s3://my-bucket/ \
  --sse AES256
```

✅ Q83: You want to restrict access to an S3 bucket to only EC2 instances with a specific role. What do you use?

A. Bucket Policy with Principal: EC2 ARN
B. IAM Group
C. VPC Endpoint
D. Resource Access Manager

Correct Answer: A. Bucket Policy with Principal: EC2 role ARN

Example:

```
{
  "Effect": "Allow",
  "Principal": {
    "AWS": "arn:aws:iam::123456789012:role/EC2S3ReadRole"
```

```
  },
  "Action": "s3:GetObject",
  "Resource": "arn:aws:s3:::my-bucket/*"
}
```

✅ **Q84: What AWS service allows querying structured data in S3 using SQL?**

A. Athena
B. Redshift
C. RDS
D. EMR

Correct Answer: A. Athena

Example Query in Athena:

```
SELECT * FROM my_bucket_data
WHERE region = 'us-east-1';
```

✅ **Q85: How would you automate Lambda deployment?**

A. Use Route 53
B. Use CodePipeline + CodeDeploy
C. Use EC2 User Data
D. Use IAM AssumeRole

Correct Answer: B. Use CodePipeline + CodeDeploy

✅ **Q86: Which CLI command gives you a list of running EC2 instances?**

A. aws ec2 run-instances
B. aws ec2 describe-instances
C. aws ec2 reboot-instances
D. aws ec2 get-instances

Correct Answer: B

```
aws ec2 describe-instances --query "Reservations[*].Instances[*].[InstanceId,State.Name]" --output table
```

✅ **Q87: How do you allow a Lambda to access DynamoDB?**

Answer:

1. Create an IAM Role with this policy:

```
{
 "Effect": "Allow",
 "Action": [
  "dynamodb:GetItem",
  "dynamodb:PutItem"
 ],
 "Resource": "arn:aws:dynamodb:us-east-1:123456789012:table/MyTable"
}
```

2. Attach the role to your Lambda function.

✅ **Q88: You want to route 80% of traffic to one Lambda and 20% to another. What do you use?**

Correct Answer: Use **Lambda Aliases + Traffic Shifting**

Example using AWS CLI:

```
aws lambda update-alias \
 --function-name MyFunction \
 --name PROD \
 --routing-config '{"AdditionalVersionWeights":{"2":0.2}}'
```

✅ **Q89: What tool lets you track IAM policy changes?**

Correct Answer: AWS CloudTrail

Example CLI to get last 10 IAM policy events:

```
aws cloudtrail lookup-events \
 --lookup-attributes AttributeKey=EventName,AttributeValue=PutUserPolicy \
 --max-results 10
```

✅ **Q90: You need to auto-tag resources based on user identity. What service helps?**

Correct Answer: AWS Lambda + CloudTrail EventBridge rule

Workflow:

- CloudTrail logs event
- EventBridge triggers Lambda

- Lambda adds tags via create-tags

✅ **Q91: How do you get a cost report grouped by service?**

```
aws ce get-cost-and-usage \
 --time-period Start=2024-06-01,End=2024-06-30 \
 --granularity MONTHLY \
 --metrics "UnblendedCost" \
 --group-by Type=DIMENSION,Key=SERVICE
```

✅ **Q92: How can you control which regions a developer can launch resources in?**

Correct Answer: Use an SCP (Service Control Policy)

Example SCP:

```
{
 "Effect": "Deny",
 "Action": "*",
 "Resource": "*",
 "Condition": {
  "StringNotEquals": {
   "aws:RequestedRegion": ["us-east-1", "us-west-2"]
  }
 }
}
```

✅ **Q93: How do you store docker images for ECS?**

Correct Answer: Amazon Elastic Container Registry (ECR)

Example Push to ECR:

aws ecr get-login-password | docker login --username AWS --password-stdin <account>.dkr.ecr.us-east-1.amazonaws.com

docker tag my-app:latest <account>.dkr.ecr.us-east-1.amazonaws.com/my-app

docker push <account>.dkr.ecr.us-east-1.amazonaws.com/my-app

✅ **Q94: How can you ensure EC2 volumes are encrypted by default?**

```
aws ec2 enable-ebs-encryption-by-default
```

✅ **Q95: What's the best way to run Spark jobs on AWS without managing servers?**

Correct Answer: AWS Glue or EMR Serverless

```
aws glue start-job-run --job-name my-spark-job
```

✅ **Q96: You want to provision resources automatically using code. What do you use?**

Correct Answer: AWS CloudFormation or CDK

CloudFormation Template Example (YAML):

```
Resources:
  MyBucket:
    Type: AWS::S3::Bucket
    Properties:
      BucketName: my-example-bucket
```

✅ **Q97: Which AWS CLI command enables versioning on an S3 bucket?**

```
aws s3api put-bucket-versioning \
  --bucket my-bucket \
  --versioning-configuration Status=Enabled
```

✅ **Q98: What tool would you use to automate infrastructure testing?**

Correct Answer: AWS CloudFormation + TaskCat or **Terraform + Terratest**

✅ **Q99: What command deploys a Lambda function from a ZIP file?**

```
aws lambda create-function \
  --function-name myFunction \
  --runtime python3.9 \
  --role arn:aws:iam::123456789012:role/lambda-role \
  --handler lambda_function.lambda_handler \
  --zip-file fileb://function.zip
```

✅ **Q100: You want to give temporary access to an S3 object. What do you use?**

Correct Answer: Presigned URL

Generate via CLI:

aws s3 presign s3://my-bucket/file.txt --expires-in 3600

☑️ Real-World Scenario: Building a Source-to-Target Data Pipeline in AWS

🎯 **Use Case: E-commerce Analytics Platform**

You are a Data Engineer working for an e-commerce company. The business wants to analyze:

- Daily order trends
- Customer behavior
- Payment and delivery performance across multiple departments and regions.

You need to build a **daily data pipeline** to ingest raw data from multiple sources and deliver it in an analytics-ready format for dashboards in Amazon QuickSight or Redshift.

🔄 **Pipeline Overview**

Stage	AWS Service	Description
Source	Amazon RDS / MySQL, CSVs in S3	Raw transactional data
Ingestion	AWS Glue Jobs or DMS	Extract data
Staging	Amazon S3	Landing zone for raw data
Processing	AWS Glue (ETL) or PySpark on EMR	Data transformation
Storage	Amazon Redshift / S3 Data Lake	Target system
Visualization	Amazon QuickSight	BI dashboards

🔧 **Step-by-Step Breakdown**

✅ **Step 1: Source Data**

- Source 1: **Amazon RDS for MySQL** → orders, customers, payments tables
- Source 2: **CSV files from store managers**, manually uploaded to s3://raw-customer-surveys/

✅ **Step 2: Ingestion to S3 (Staging Zone)**

Option 1: Using AWS Glue for JDBC Ingestion

```
# Python Glue job snippet
datasource = glueContext.create_dynamic_frame.from_options(
   connection_type="mysql",
   connection_options={"url": "jdbc:mysql://rds-endpoint", "user": "admin", "password": "******"},
   table_name="orders"
)
datasink = glueContext.write_dynamic_frame.from_options(
   frame=datasource,
   connection_type="s3",
   connection_options={"path": "s3://ecommerce-data/raw/orders/"},
   format="parquet"
)
```


Option 2: Use DMS for continuous ingestion (CDC)

- Set up a replication task from RDS → S3 in Parquet/CSV format
- Enables **near real-time** ingestion

✅ Step 3: Data Transformation

You use **AWS Glue PySpark ETL jobs** to:

- Clean nulls
- Join orders + customers
- Convert currencies
- Flatten nested fields

```
# Glue PySpark join example

orders_df = glueContext.create_dynamic_frame.from_catalog(database="ecomm", table_name="orders").toDF()

cust_df = glueContext.create_dynamic_frame.from_catalog(database="ecomm", table_name="customers").toDF()


result_df = orders_df.join(cust_df, orders_df.customer_id == cust_df.id, "left") \

   .withColumn("order_month", F.date_format("order_date", "yyyy-MM"))


result_df.write.format("parquet").save("s3://ecommerce-data/processed/orders/")
```

✅ Step 4: Load to Target

Target 1: Amazon Redshift (for analytics and dashboards)
Target 2: Amazon S3 Data Lake (for data science team)

Glue job loads processed Parquet files into Redshift:

```
COPY orders

FROM 's3://ecommerce-data/processed/orders/'

IAM_ROLE 'arn:aws:iam::123456789012:role/MyRedshiftRole'

FORMAT AS PARQUET;
```

✅ Step 5: Visualization in QuickSight

- Connect Amazon QuickSight to Redshift or Athena
- Create dashboards for:
    - Total orders by region
    - Delay in payment vs delivery
    - Abandoned carts

📌 Optional Enhancements

Feature	AWS Service Used
Data Validation	AWS Deequ + Glue
Orchestration	AWS Step Functions or Managed Workflows for Apache Airflow
Monitoring & Alerts	CloudWatch + SNS
Cataloging	AWS Glue Data Catalog
Versioning	Lake Formation + S3
Access Control	Lake Formation + IAM

💼 Interview Talking Points

If asked about a pipeline in an interview, you can say:

In a recent e-commerce data project, I built a Glue-based pipeline to ingest and process transactional data from RDS and CSVs, applied PySpark transformations for cleansing and enrichment, and loaded the data into Redshift and S3. I used Glue Catalog for schema management and QuickSight for visualization. I also added error handling via Step Functions and alerting using CloudWatch.