PREPKLOUD FIELD GUIDE SERIES

PrepKloud

AWS Solutions Architect Associate The Complete SAA-C03 Field Guide

A practical, scenario-first guide to designing AWS architectures that are secure, resilient, high-performing, and cost-optimized.

1st Edition · 2026 · prepkloud.com

About this book

AWS Solutions Architect Associate: The Complete SAA-C03 Field Guide is published by PrepKloud, an independent cloud-certification preparation platform. It is aligned to the current SAA-C03 four-domain outline: Design Secure Architectures (30%), Design Resilient Architectures (26%), Design High-Performing Architectures (24%), and Design Cost-Optimized Architectures (20%).

This guide teaches architecture judgment rather than a disconnected catalog of AWS services. Every chapter starts with requirements, constraints, and tradeoffs—the way the SAA-C03 exam frames its scenarios—then maps those decisions to AWS services and design patterns.

How to use this book: read Chapters 1–3 first, then work through the four weighted domains. Do the hands-on architecture exercises in Chapter 14 using a disposable account, synthetic data, least privilege, and a cleanup checklist. Attempt every original practice question before checking its explanation.

Disclaimer: AWS, Amazon EC2, Amazon S3, AWS Lambda, and SAA-C03 are trademarks of Amazon Web Services, Inc. This independent guide is not endorsed by or affiliated with AWS. AWS services, limits, exam content, and pricing can change; confirm current details against the official AWS Certification exam guide before scheduling. This book contains no leaked, recalled, or exam-dump material.

Table of Contents

CHAPTER 01

How SAA-C03 Tests Architecture Judgment

SAA-C03 does not reward remembering the longest list of AWS services. It rewards choosing the simplest architecture that satisfies every stated requirement: security, availability, performance, operational effort, and cost. Most questions deliberately give you more than one plausible service; the correct answer is the one that best matches the constraints without adding unnecessary complexity.

Domain	Weight	Primary question

Design Secure Architectures	30%	Who can access which resource and data, through what controlled path?
Design Resilient Architectures	26%	How does the solution tolerate failure, preserve work, and recover?
Design High-Performing Architectures	24%	How does it meet latency, throughput, and scaling requirements?
Design Cost-Optimized Architectures	20%	How does it meet requirements without paying for unused capacity?

> **Exam focus:** Treat phrases such as **most cost-effective**, **least operational overhead**, **highly available**, and **lowest latency** as mandatory design constraints—not filler.

A five-step question method

1. Underline the required outcome and every non-negotiable constraint.
2. Identify the workload type: web, batch, event-driven, database, analytics, hybrid, or migration.
3. Eliminate answers that violate an explicit constraint.
4. Compare the remaining answers on managed-service fit and operational overhead.
5. Select the simplest AWS-native design that meets all stated requirements.

> **Exam trap:** "Use the biggest instance" and "add every possible security service" are not architecture answers. The exam regularly penalizes over-engineering when a managed, scoped service meets the need.

Chapter 1 checklist

☐ I know the four domains and their relative weights.

☐ I read requirements before evaluating services.

☐ I can explain why the simplest managed solution is often preferred.

CHAPTER 02

The AWS Architecture Decision Framework

Architecture choices become easier when you turn vague business statements into measurable constraints. “The site must never go down” usually means multi-AZ availability and tested recovery; “customers worldwide need fast downloads” means edge caching and regional placement; “the team is small” means a managed service has more value than self-managed servers.

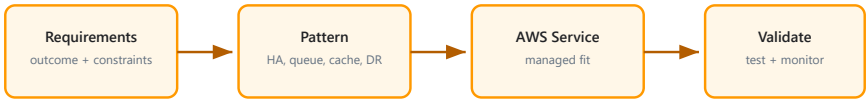

Figure 2.1 — Start with requirements, then choose a pattern and service.

If the scenario says...	Think first about...
“Variable traffic”	Auto Scaling, serverless, buffering, on-demand capacity.
“No data loss”	Durability, replication mode, backups, RPO/RTO.
“Global users”	CloudFront, Route 53 routing, multi-Region only when needed.
“No server management”	Lambda, Fargate, managed databases, managed messaging.
“Private access only”	VPC design, private subnets, endpoints, IAM, security groups.

Scenario

Spiky image-processing workload

A media site receives unpredictable batches of image uploads. Processing can happen asynchronously and the team does not want to manage servers.

Decision path: Store uploads in S3, trigger event-driven processing via SQS/Lambda or EventBridge/Lambda, and retain failed work in a dead-letter queue. This absorbs bursts, removes server administration, and isolates failures.

Chapter 2 checklist

☐ I can turn business wording into technical constraints.

☐ I can name the common pattern behind a scenario before naming AWS services.

CHAPTER 03

AWS Global Infrastructure & Shared Responsibility

A Region is a separate geographic area. Each Region contains multiple isolated Availability Zones (AZs) connected by low-latency networking. A multi-AZ design protects against an AZ failure; a multi-Region design protects against a Regional outage and can reduce global-user latency, but adds cost and operational complexity.

Scope	Best use	Exam signal
One AZ	Noncritical or development workload	Lowest cost, not highly available

Multi-AZ	High availability inside one Region	“Tolerate an Availability Zone failure”
Multi-Region	Regional disaster recovery or global latency	“Survive a Regional outage” or global active users

Shared responsibility

AWS secures the cloud infrastructure: facilities, hardware, networking, and the managed-service infrastructure. You secure what you put in the cloud: identities, permissions, data classification, encryption choices, network configuration, operating systems for EC2, and application code. The exact boundary moves with the service model—AWS manages more for Lambda than for EC2.

> **Exam focus:** If a question asks who patches the guest operating system, the answer is the customer for EC2. For a

> managed database service, AWS manages the underlying infrastructure while the customer still owns access, data, and configuration decisions.

Chapter 3 checklist

☐ I can choose multi-AZ vs multi-Region based on the failure scope.

☐ I understand the shared-responsibility boundary for EC2 and managed services.

Secure Access, Identities & Multi-Account Design

IAM decisions are not simply “create a user.” Start with temporary credentials and least privilege. Human workforce access should generally use federation/identity-center patterns; AWS workloads should use IAM roles. Roles issue short-lived credentials and remove the need to store long-lived keys in code.

Requirement	Best-fit control
Application on EC2 needs S3 access	Attach an IAM role to the instance.

External account needs narrowly scoped access	Cross-account IAM role with a trust policy.
Organization-wide permission boundary	AWS Organizations OUs with SCPs; SCPs do not grant permissions.
Central sign-in for workforce users	Federation / IAM Identity Center, with role-based access.
Secrets requiring rotation	Secrets Manager, not values embedded in code.

> **Exam trap:** An SCP is a guardrail, not a permission grant. A principal still needs an IAM identity or resource policy that permits the action.

Multi-account design

Separate production, development, security/log archive, and shared-services workloads into accounts to reduce blast radius and establish

clearer billing, audit, and permission boundaries. Use AWS Organizations for consolidated governance, SCPs for preventive boundaries, and cross-account roles for controlled access.

Scenario

Centralized security review

An enterprise needs its security team to investigate CloudTrail logs across all application accounts without sharing long-lived credentials.

Decision path: centralize logs in a dedicated log archive account and allow the security team to assume cross-account roles. This uses temporary credentials and preserves account isolation.

Chapter 4 checklist

☐ I prefer roles and temporary credentials over embedded access keys.

☐ I know what SCPs do—and what they do not do.

☐ I can explain why multi-account design reduces blast radius.

Secure Networks, Applications & Data

Network security is layered. Security groups are stateful firewalls attached to ENIs/instances; network ACLs are stateless subnet-level controls. Use security groups for the usual “allow only this traffic to this workload” design; use NACLs when a subnet-level, ordered allow/deny boundary is specifically required.

Control	Key property	Use it when…
Security group	Stateful; allow rules only	Controlling traffic to workloads.

Network ACL	Stateless; ordered allow/deny rules	Applying subnet-wide guardrails.
VPC endpoint	Private service access	Workloads must reach supported AWS services without public internet/NAT.
AWS WAF	Layer 7 web filtering	Protecting CloudFront, ALB, or API Gateway from application-layer attacks.

Encryption choices

Encryption at rest protects stored data (for example S3, EBS, RDS, and DynamoDB encryption). Encryption in transit protects traffic using TLS. AWS KMS manages encryption keys for most AWS-native use cases; CloudHSM provides dedicated hardware where strict compliance or direct cryptographic control requires it.

> **Exam focus:** A private subnet does not automatically mean “no data leaves the VPC.” For private access to services such as S3, use the appropriate VPC endpoint when the scenario requires avoiding an internet gateway or NAT path.

Chapter 5 checklist

- ☐ I know stateful security groups vs stateless NACLs.
- ☐ I distinguish encryption at rest from encryption in transit.
- ☐ I can recognize when a VPC endpoint is required.

Resilient Compute & Decoupled Workloads

Resilience starts by removing single points of failure. An application tier commonly uses an Application Load Balancer across multiple AZs with Auto Scaling groups spanning those AZs. Stateless application servers are easier to replace, scale, and distribute than servers that retain local session state.

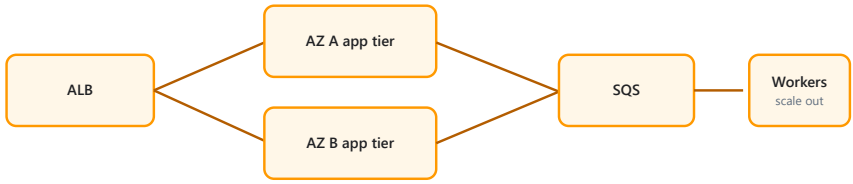

Figure 6.1 — Multi-AZ application tier plus a queue decouples request handling from asynchronous work.

Queue-based decoupling

Amazon SQS stores work until consumers can process it. It absorbs bursts, isolates failures, and lets producers and consumers scale independently. Standard queues prioritize high throughput and at-least-once delivery; FIFO queues are for strict ordering and exactly-once processing semantics. Dead-letter queues preserve repeatedly failing messages for investigation.

> **Exam trap:** Do not select a FIFO queue merely because a system is “important.” Select it only when ordering or

> deduplication constraints require it; Standard queues are the default high-throughput choice.

Chapter 6 checklist

- ☐ I can design a multi-AZ web tier with load balancing and Auto Scaling.
- ☐ I know why queues improve resilience and burst handling.
- ☐ I can choose Standard vs FIFO based on actual requirements.

Resilient Data, Backups & Disaster Recovery

Data architecture questions often hinge on two recovery measurements. **RPO** (recovery point objective) is acceptable data loss measured in time. **RTO** (recovery time objective) is acceptable time to restore service. A design must meet both, not merely “have a backup.”

Pattern	Typical cost	Recovery characteristics
Backup and restore	Lowest	Hours; suited to less critical workloads.

Pilot light	Low to medium	Core components running; scale during recovery.
Warm standby	Medium to high	Reduced-capacity environment can scale quickly.
Multi-site active/active	Highest	Fastest recovery and global availability.

Database availability patterns

RDS Multi-AZ is for high availability: synchronous standby replication and automatic failover, not read scaling. Read replicas are for read scale and can support regional-disaster-recovery patterns; they replicate asynchronously. Aurora combines managed relational compatibility with distributed storage and options for read scaling and multi-Region patterns.

> **Exam focus:** “Improve read performance” points to read replicas. “Automatically fail over when the primary database instance fails” points to Multi-AZ. These are not interchangeable.

S3 resilience

Amazon S3 is designed for high durability. Use versioning to protect against accidental overwrite/delete, lifecycle policies to transition or expire objects, and replication where a business requirement calls for copies in another Region or account. Remember that versioning helps recover deleted objects but must be paired with lifecycle rules if old versions need cost control.

Chapter 7 checklist

- ☐ I can explain RPO vs RTO.

- [ ] I distinguish Multi-AZ from read replicas.
- [ ] I can match DR patterns to recovery speed and cost.

High-Performing Compute, Storage & Databases

Performance questions usually tell you the bottleneck: CPU, memory, latency, IOPS, throughput, reads, writes, or concurrent connections. Select the resource model that addresses that bottleneck—do not pick based on a service being familiar.

Need	Architecture direction
Short, event-driven computation	Lambda; consider concurrency and timeout constraints.
Long-running containerized work	ECS/Fargate or EKS, depending on operational needs.

Predictable VM baseline	EC2 with Savings Plans/Reserved Instances where appropriate.
High IOPS database storage	Provisioned IOPS EBS volumes or managed database options.
Shared Linux file system across instances	EFS.
Massively scalable key-value access	DynamoDB, designed around access patterns.

Caching

Caching reduces repeated database or origin work. CloudFront caches content near users; ElastiCache accelerates application data access; API Gateway caching can reduce backend calls. Cache only data that can tolerate its freshness model, and design a clear invalidation/TTL approach.

Scenario

Global product catalog

Users around the world request the same product images and mostly read-only catalog pages. The origin is overloaded during marketing events.

Decision path: place CloudFront in front of the origin, configure cache behavior and HTTPS, and store static assets in S3 or serve dynamic pages through an origin. This reduces global latency and origin load simultaneously.

Chapter 8 checklist

- [ ] I can identify a stated performance bottleneck before choosing a service.
- [ ] I know the difference between edge caching and in-memory application caching.
- [ ] I can match storage and database types to their access patterns.

Networking, DNS & Content Delivery

Networking scenarios are about traffic path and scope. Inside a VPC, subnets define placement; route tables define paths; security groups/NACLs define filtering. Between networks, use VPC peering for a simple, non-transitive connection, and Transit Gateway for hub-and-spoke, transitive connectivity across many VPCs, VPNs, or Direct Connect connections.

Service	Use it when...
Application Load Balancer	HTTP/HTTPS, host/path-based routing, WebSockets, Layer 7 decisions.

Network Load Balancer	TCP/UDP/TLS, static IPs, very high performance, Layer 4 traffic.
Route 53 failover routing	Active-passive DNS routing based on health checks.
Route 53 latency routing	Route users to the region that gives lowest latency.
CloudFront	Cache and deliver content from edge locations globally.

> **Exam trap:** VPC peering is not transitive. If VPC A peers with B and B peers with C, A cannot automatically route to C through B. Use Transit Gateway for that hub-and-spoke requirement.

Private hybrid connectivity

AWS Direct Connect provides dedicated private connectivity to AWS for steady high-bandwidth or predictable-latency hybrid requirements. Site-to-

Site VPN can be deployed faster over the internet and is often used as a backup or lower-cost alternative. The correct answer depends on bandwidth, consistency, lead time, and confidentiality requirements.

Chapter 9 checklist

- ☐ I choose ALB vs NLB by protocol/routing layer.
- ☐ I know when Transit Gateway is needed instead of peering.
- ☐ I can match Route 53 routing policy to a requirement.

Cost-Optimized Architectures

Cost optimization is not “choose the cheapest service.” It is delivering the required outcome at the lowest total cost, including engineering effort, operational risk, data transfer, and unused capacity. Start with utilization and access frequency.

Cost pattern	Decision
Steady, predictable EC2 use	Consider Savings Plans or Reserved Instances for the baseline.
Interruptible batch/flexible workloads	Use Spot Instances with interruption-tolerant design.
Unpredictable request volume	On-demand/serverless or Auto Scaling rather than idle capacity.

Rarely accessed S3 data	Lifecycle transition to an appropriate infrequent-access/archive class.
Repeated global content delivery	CloudFront can reduce origin load and transfer costs.

Right sizing and visibility

Use CloudWatch metrics and cost tools to identify underused resources. A correct design also labels and groups resources for allocation, uses budgets/alerts to detect surprises, and enables lifecycle policies where data age predicts access frequency.

> **Exam focus:** Spot is not the answer for a database primary or an uninterruptible production service. It is the answer when the workload can handle interruptions, checkpoint progress, or draw work from a queue.

Chapter 10 checklist

☐ I choose pricing model based on workload predictability and interruption tolerance.

☐ I can use S3 lifecycle rules to align storage cost with access frequency.

☐ I know cost optimization includes visibility, rightsizing, and managed-service tradeoffs.

CHAPTER 11

Migration, Hybrid & Data Transfer Decisions

Migration questions are usually about volume, time, bandwidth, and whether the transfer is one-time or ongoing. DataSync is a managed online transfer service for ongoing or repeatable movement. The AWS Snow family is designed for large offline transfers when the network is insufficient. Storage Gateway provides hybrid file, volume, or tape patterns that connect on-premises environments to AWS storage.

Requirement	Best direction

Ongoing online file transfer/synchronization	AWS DataSync
Petabyte-scale, constrained network, one-time seed	AWS Snow family
Hybrid NFS/SMB interface backed by S3	Storage Gateway File Gateway
Dedicated enterprise connectivity	Direct Connect, often with VPN backup
Database migration with minimal downtime	AWS Database Migration Service pattern

> **Exam focus:** Do not select physical transfer merely because the data set is “large.” The question must indicate that network transfer is impractical relative to timeline/bandwidth or that offline transfer is specifically required.

Chapter 11 checklist

- [ ] I can distinguish online sync from offline bulk transfer.
- [ ] I can map hybrid storage interface requirements to Storage Gateway patterns.

CHAPTER 12

Observability, Governance & Infrastructure as Code

Production architecture requires evidence. CloudWatch provides metrics, logs, dashboards, and alarms. CloudTrail records API activity—who performed which action, from where, and when. AWS Config records resource configuration state and evaluates compliance. They are complementary: CloudTrail answers "who changed it?"; Config answers "what is the resource configuration and how did it change?"

Service	Primary question it answers

CloudWatch	Is the system healthy and performing as expected?
CloudTrail	Who called which AWS API action and when?
AWS Config	What is the resource configuration; is it compliant?
CloudFormation	How do we deploy repeatable AWS infrastructure?
AWS Backup	How do we centrally manage backup plans and retention?

Infrastructure as code (IaC) turns architecture into reviewable, repeatable templates. CloudFormation change sets preview updates, drift detection identifies manual deviation, and StackSets support multi-account/multi-Region rollout patterns.

Chapter 12 checklist

- [ ] I distinguish CloudWatch, CloudTrail, and Config.
- [ ] I can explain why IaC improves repeatability and reviewability.

CHAPTER 13

Architecture Scenario Patterns

Pattern 1

Highly available web application

Requirements: public HTTPS app, variable traffic, AZ-failure tolerance, relational database, minimal operations.

Pattern: Route 53 → ALB across AZs → Auto Scaling EC2 or ECS/Fargate app tier → RDS Multi-AZ. Store static assets in S3 and place CloudFront in front where global caching is useful.

Pattern 2

Asynchronous order processing

Requirements: bursts of orders, no lost work, workers may fail/retry, no ordering requirement.

Pattern: application publishes to SQS Standard; stateless workers consume and scale; failed messages route to a DLQ. Use idempotent consumers because Standard delivery is at least once.

Pattern 3

Protected multi-account platform

Requirements: isolated environments, centralized governance and audit, no shared access keys.

Pattern: Organizations OUs/account boundaries + SCP guardrails + central log archive + IAM Identity Center/federation + cross-account roles for controlled administration.

Pattern 4

Cost-sensitive analytics archive

Requirements: retain years of logs, rare access, ad hoc queries when requested.

Pattern: S3 with lifecycle transitions, partitioned/compressed data where practical, and Athena for serverless SQL queries directly against S3 rather than permanently running an analytics cluster.

Chapter 13 checklist

☐ I can sketch four common SAA-C03 architecture patterns from memory.

☐ I can explain the requirement that justifies every component in each pattern.

CHAPTER 14

Hands-On Lab & Portfolio Blueprint

Build a small, disposable architecture that demonstrates security, resilience, performance, and cost choices rather than trying to deploy every AWS service. Use a dedicated sandbox account or tightly scoped temporary environment; never use production data or commit secrets.

1. Write a one-page architecture brief: application goal, traffic assumptions, availability target, RTO/RPO, and monthly budget ceiling.

2. Diagram a two-AZ web tier: public load balancer, private application subnets, security groups, and a managed data layer.
3. Deploy infrastructure through CloudFormation or another approved IaC workflow; record parameters and expected outputs.
4. Add CloudWatch alarms, CloudTrail logging, and tags for cost allocation.
5. Run one controlled failure test (for example, remove an app instance), capture the recovery evidence, then clean up every temporary resource.

Portfolio evidence	What it proves
Architecture diagram	You can explain traffic flow and trust boundaries.

IaC template + change notes	You can create repeatable infrastructure.
Failure test result	You tested resilience rather than assuming it.
Cost estimate + cleanup proof	You consider operational and financial accountability.

> **Study tip:** The strongest interview explanation is: requirement → decision → tradeoff → validation. Example: "We used SQS to absorb bursts, accepted at-least-once delivery, made consumers idempotent, and verified retry/DLQ behavior."

Chapter 14 checklist

☐ I have a safe, disposable lab plan.

☐ I will collect architecture, test, and cleanup evidence.

- [ ] I can explain each decision as a tradeoff.

CHAPTER 15

40 Practice Questions with Explanations

These are original study questions designed to train decision-making against the SAA-C03 domains. They are not recalled or leaked exam content.

Security & Resilience (Q1–Q20)

1. An EC2 application needs read-only access to one S3 bucket. What is the most secure credential approach?

Answer: Attach an IAM role with least-privilege S3 permissions to the EC2 instance.

Roles provide temporary credentials and avoid embedded long-lived keys.

2. Which control governs the maximum permissions available to accounts in an organizational unit?

Answer: A service control policy (SCP).

SCPs set permission boundaries but do not grant permissions themselves.

3. A workload in a private subnet must access S3 without routing through the public internet or a NAT gateway. What should be used?

Answer: An S3 VPC endpoint.

It provides private connectivity to the supported service.

4. Which firewall is stateful and attached to an instance/network interface?

Answer: A security group.

NACLs are stateless and operate at the subnet level.

5. A relational database must automatically fail over if its primary instance fails. What pattern best fits?

Answer: RDS Multi-AZ.

Multi-AZ is for high availability, not read scaling.

6. An app requires additional read capacity from a relational database. What should be considered?

Answer: Read replicas.

Read replicas offload read traffic; they are not the same as Multi-AZ standby.

7. A message-processing system requires strict ordering within a business entity and deduplication semantics. Which queue type fits?

Answer: SQS FIFO.

FIFO is selected for ordering/exactly-once processing semantics, not general importance.

8. What preserves messages that repeatedly fail processing for later investigation?

Answer: A dead-letter queue.

It avoids losing or endlessly retrying poisoned messages.

9. What is the RPO?

Answer: The maximum acceptable data loss measured in time.

RTO measures time to restore service.

10. A company needs recovery from a complete Region outage with a low RTO and can afford continuously running secondary capacity. Which DR strategy is closest?

Answer: Warm standby or active/active, depending on the stated recovery target.

Backup/restore is cheaper but slower; active/active costs the most and recovers fastest.

11. What S3 feature best protects against accidental object overwrite or deletion?

Answer: Versioning.

Pair with lifecycle configuration to control old-version cost.

12. Which service is a better fit for a secret that needs automatic rotation?

Answer: AWS Secrets Manager.

It is purpose-built for managed secret rotation integrations.

13. A public application must block common web exploits at the HTTP layer. Which service should be considered?

Answer: AWS WAF.

It protects Layer 7 endpoints such as CloudFront, ALB, and API Gateway.

14. Which AWS service records API activity such as who changed a security group?

Answer: AWS CloudTrail.

CloudTrail records API calls; Config records resource configuration state.

15. How can a security team access multiple application accounts without sharing permanent credentials?

Answer: Assume scoped cross-account IAM roles.

This creates temporary, auditable access.

16. A workload must continue during loss of one Availability Zone. What is the minimum infrastructure scope?

Answer: Deploy across at least two Availability Zones.

A single-AZ deployment remains an AZ-level single point of failure.

17. Which design makes a web tier easier to replace and Auto Scale?

Answer: Keep the application tier stateless and externalize session/state data.

Instances can be replaced without losing user state.

18. Which service centrally automates backup plans and retention across supported AWS services?

Answer: AWS Backup.

It provides centralized policy-driven backup administration.

19. A private connection is required between on-premises infrastructure and AWS with predictable high bandwidth. Which choice best fits?

Answer: AWS Direct Connect.

VPN is internet-based; Direct Connect is a dedicated connection.

20. What is the common use of a Site-to-Site VPN alongside Direct Connect?

Answer: Backup connectivity or a faster-to-provision hybrid path.

It can complement a dedicated connection for resilience.

Performance & Cost (Q21–Q40)

21. Global users request the same static images and the origin is overloaded. What service should be added first?

Answer: Amazon CloudFront.

An edge cache lowers user latency and reduces origin load.

22. Which load balancer supports HTTP path-based routing?

Answer: Application Load Balancer.

ALB is Layer 7; NLB operates at Layer 4.

23. Which load balancer is the best fit for TCP/UDP workloads requiring static IP addresses?

Answer: Network Load Balancer.

NLB is designed for high-performance Layer 4 traffic.

24. A company needs hub-and-spoke routing across dozens of VPCs. What service should it choose?

Answer: AWS Transit Gateway.

VPC peering is point-to-point and non-transitive.

25. Which Route 53 routing policy supports active-passive failover using health checks?

Answer: Failover routing.

Latency routing chooses lowest-latency resources instead.

26. An API experiences unpredictable traffic with long idle periods. Which compute pricing/operation model commonly fits?

Answer: Serverless/on-demand, such as Lambda when its constraints fit.

It avoids paying for idle servers and scales with requests.

27. A batch workload can checkpoint work and tolerate interruption. How can compute cost be reduced most aggressively?

Answer: Use Spot Instances.

Spot is appropriate only for interruption-tolerant workloads.

28. A production service uses a stable baseline of EC2 capacity year-round. What option can reduce cost?

Answer: Savings Plans or Reserved Instances for the predictable baseline.

Keep flexibility for variable capacity with On-Demand/Auto Scaling.

29. Which service provides a shared, elastic Linux file system for multiple EC2 instances?

Answer: Amazon EFS.

EFS is a managed NFS-style shared file system.

30. A workload needs single-digit millisecond key-value reads at massive scale. Which database is often appropriate?

Answer: Amazon DynamoDB.

It is a managed NoSQL service designed around access patterns and scale.

31. A team needs ad hoc SQL queries against data in S3 without operating a cluster. Which service fits?

Answer: Amazon Athena.

Athena is serverless SQL over data stored in S3.

32. Which S3 feature automatically transitions old objects to lower-cost classes or expires them?

Answer: S3 lifecycle rules.

Use them when access frequency changes predictably over time.

33. What should be checked first when an architecture's bill increases unexpectedly?

Answer: Utilization and cost-allocation data, then identify idle/oversized resources or transfer growth.

Do not blindly change services before identifying the driver.

34. Which managed cache is commonly used to reduce repeated database reads for application data?

Answer: Amazon ElastiCache.

CloudFront caches edge-delivered content; ElastiCache is application-level in-memory caching.

35. A data transfer is too large for the available network timeline and only needed once. What is the likely AWS direction?

Answer: AWS Snow family.

Physical devices suit offline bulk transfer where online transfer is impractical.

36. An organization needs repeatable, version-controlled infrastructure deployment. Which service provides native templates and change sets?

Answer: AWS CloudFormation.

Change sets preview updates before execution.

37. Which service records configuration history and evaluates resource compliance?

Answer: AWS Config.

It complements CloudTrail's API activity history.

38. How can a company direct users to the AWS Region with the lowest network latency?

Answer: Route 53 latency-based routing.

It directs based on latency measurements, not a fixed traffic percentage.

39. What design avoids managing servers for a multi-step workflow that can take longer than a Lambda invocation?

Answer: AWS Step Functions orchestrating suitable managed services.

It coordinates long-running, stateful workflows with retries and branching.

40. A solution must use the least operational effort while keeping worker capacity aligned with a queue backlog. What principle applies?

Answer: Use managed, autoscaling consumers where suitable and scale based on queue demand.

The best answer depends on workload constraints, but managed autoscaling reduces operational overhead.

Chapter 15 checklist

☐ I attempted every question before reading the answer.

☐ I can explain the rejected alternative, not just the correct choice.

CHAPTER 16

6-Week Study Planner & Cheat Sheet

Week	Study focus	Applied work
1	Ch. 1–3: exam method, infrastructure, shared responsibility	Sketch Region/AZ designs and the responsibility boundary.
2	Ch. 4–5: IAM, organizations, networks, encryption	Map identities, trust policies, and private traffic paths.
3	Ch. 6–7: compute resilience, queues, data/DR	Design an RPO/RTO matrix and queue consumer flow.
4	Ch. 8–9: performance, databases, networking/CDN	Compare ALB/NLB, EFS/EBS, peering/Transit Gateway.

5	Ch. 10–12: cost, migration, observability/IaC	Estimate a lab and define monitoring/cleanup evidence.
6	Ch. 13–15: scenarios and timed review	Do practice questions twice and remediate weak domains.

Remember these distinctions

- Multi-AZ = HA; read replica = read scaling.
- Security group = stateful; NACL = stateless.
- CloudTrail = API activity; Config =

Exam language → pattern

- “Least operations” → managed/serverless.
- “Burst absorption” → queue/event-driven.
- “Global low latency” → CloudFront/Route 53.

configuration state.

- ALB = Layer 7; NLB = Layer 4.
- Peering = non-transitive; Transit Gateway = hub-and-spoke.
- “No lost messages” → durable queue + retries/DLQ.
- “Predictable baseline” → Savings Plan/Reserved capacity.

CHAPTER 17

Glossary, Exam-Day Strategy & FAQ

Availability Zone (AZ)

An isolated location within an AWS Region designed to limit failure scope.

CloudTrail

A record of AWS API activity for audit, security investigation, and troubleshooting.

Dead-letter queue (DLQ)

A queue for messages that repeatedly fail normal processing.

IAM role

An identity with permissions that can be assumed to obtain temporary credentials.

RPO

Maximum acceptable data loss measured in time.

RTO

Maximum acceptable time to restore service after disruption.

SCP

An Organizations guardrail that limits maximum permissions in accounts/OUs; it does not grant access.

VPC endpoint

Private connectivity from a VPC to supported AWS services without public internet routing.

Exam-day strategy

- Read the final sentence first: it often states the exact requirement being tested.
- Reject solutions that fail a stated constraint before comparing services.
- Prefer managed AWS-native patterns when they meet all requirements with less operational work.
- For a two-answer question, verify that each selected answer solves a different required part of the design.

Is SAA-C03 beginner-friendly?

Answer: It is Associate-level and assumes practical AWS architecture judgment.

You do not need every AWS service memorized, but you should be comfortable comparing design tradeoffs and reading scenario requirements.

Is Cloud Practitioner required first?

Answer: No.

It may help complete beginners, but AWS does not require it before SAA-C03.

How should I use practice questions?

Answer: Diagnose a decision gap, then return to the underlying scenario pattern.

Memorizing answer text is weaker than being able to explain the requirement, tradeoff, and rejected alternative.

Thank you for studying with PrepKloud. Visit prepkloud.com for the SAA-C03 roadmap, original practice questions, flashcards, and hands-on projects.