**Top 50 Tough Scenario Based questions on AWS related topics (Difficult Interview Preparation)**

If you prepare below questions, you can crack any high profile AWS interview easily

**Cost-Saving & Performance Optimization**

**Scenario 1**: You need to optimize a highly transactional system hosted on AWS to reduce costs while improving performance. Describe a cost-effective strategy to move from Provisioned IOPS SSD (io1) to General Purpose SSD (gp3) volumes. How would you ensure minimal performance degradation during this transition?

**Scenario 2**: Your application uses several EC2 instances that exhibit high CPU and low memory utilization. How would you approach resizing the instances to balance performance and cost? Which AWS tools or strategies would you use to make data-driven decisions?

**Scenario 3**: You’re tasked with reducing costs in an auto-scaling group where the workload has unpredictable spikes. How would you redesign the auto-scaling policies to optimize both cost and performance?

**Scenario 4**: During performance testing, you notice a bottleneck caused by a third-party API call. What strategies would you use to cache the API response and reduce costs? Explain how you would balance caching costs with API request savings.

**Scenario 5**: A team requests a recommendation to reduce the cost of large-scale video transcoding. Would you suggest a serverless solution (like AWS Lambda) or spot instances? Justify your answer in terms of both cost and performance.

**Caching**

**Scenario 6**: You’re tasked with optimizing read-heavy database workloads with a caching layer. How would you decide between Redis and Memcached, considering both cost and performance?

**Scenario 7**: A web application uses an in-memory cache (Redis) extensively. How would you tune the eviction policies and TTL settings to ensure that the cache provides optimal cost-to-performance benefits during traffic spikes?

**Scenario 8**: An API gateway is experiencing frequent cache misses, leading to increased response times. What strategies would you implement to minimize cache misses while keeping the cost under control?

**Scenario 9**: You’re managing a distributed cache across multiple regions to serve a global user base. What are some cost-effective ways to minimize data replication lag while maximizing cache performance?

**Scenario 10**: A business-critical application relies heavily on database queries. How would you architect a cost-effective read-through or write-through cache solution that doesn’t compromise the performance or availability of the database?

**Database Optimization**

**Scenario 11**: Your organization is migrating from an on-premises SQL database to a cloud-based managed database (RDS). What strategies would you implement to optimize performance while keeping costs under control?

**Scenario 12**: A mission-critical database has unpredictable traffic patterns and fluctuating workloads. How would you design a scaling strategy for Aurora (or another cloud-native database) to handle such traffic efficiently while minimizing costs?

**Scenario 13**: You’re tasked with reducing the storage costs of a database containing 10 years of historical data, which is queried rarely. Describe your approach to partitioning or archiving this data while keeping query performance acceptable.

**Scenario 14**: Your relational database is suffering from slow queries due to poorly designed indexes. How would you systematically identify and optimize the most expensive queries without driving up operational costs?

**Scenario 15**: A NoSQL database workload has heavy write activity, and you’re experiencing bottlenecks. What would be your approach to optimize both the performance and the cost of the NoSQL database?

**Networking**

**Scenario 16**: A global application suffers from high latency between users in different regions. How would you use a combination of CDN and Anycast routing to reduce latency while controlling the costs associated with these services?

**Scenario 17**: Your application uses multiple VPCs in different AWS regions. How would you optimize the cross-region data transfer costs while ensuring low latency and high availability?

**Scenario 18**: A large-scale microservices architecture is seeing significant costs from inter-service communication due to network traffic. How would you approach minimizing these costs while maintaining performance and reliability?

**Scenario 19**: You’re tasked with optimizing the costs for a system that uses large amounts of network bandwidth for data replication between regions. What techniques or AWS services would you recommend to minimize data transfer costs?

**Scenario 20**: Your company’s hybrid cloud setup uses both AWS and on-premises infrastructure. Describe a strategy to minimize the cost of data transfer between the on-prem network and the cloud.

**Cloud Resource Utilization**

**Scenario 21**: You’re responsible for identifying underutilized cloud resources in a large enterprise. What steps would you take to efficiently identify and address these resources without negatively affecting performance?

**Scenario 22**: Some resources in your AWS environment appear to be over-provisioned. Describe how you would use AWS Cost Explorer, Trusted Advisor, or similar tools to optimize these resources without compromising performance.

**Scenario 23**: Write a script that helps identify under-utilized EC2 instances across multiple regions. How would you ensure that the script takes into account utilization patterns to avoid prematurely decommissioning critical resources?

**Scenario 24**: A team reports that their reserved instances aren’t being fully utilized. How would you design a strategy to shift workloads to maximize the use of reserved instances without affecting performance?

**Scenario 25**: For a Kubernetes cluster running on AWS, how would you balance the use of spot instances and on-demand instances to minimize costs while maintaining high availability?

**Cost Control in Cloud Environments**

**Scenario 26**: Your company’s cloud bill has been rising, and management has asked for a 20% reduction. How would you audit the infrastructure to achieve this reduction without affecting performance?

**Scenario 27**: You’re tasked with creating a highly available system across multiple AWS regions. What strategies would you use to minimize cross-region replication and data transfer costs while maintaining reliability?

**Scenario 28**: A data processing pipeline in your environment is over-provisioned for the task it handles, but it needs to handle occasional large spikes. How would you optimize its resource usage and scale dynamically to avoid unnecessary costs?

**Scenario 29**: You need to reduce the costs of running a fleet of EC2 instances but still handle unpredictable workloads. What combination of reserved instances, spot instances, and savings plans would you recommend, and how would you manage the risks?

**Scenario 30**: An organization uses a cloud-native service (e.g., AWS Lambda) for a compute-intensive task that results in unexpectedly high costs. How would you profile the workload and suggest a more cost-efficient architecture?

**Monitoring and Cost Management]**

**Scenario 31**: You’re responsible for implementing cost and performance monitoring for a multi-cloud environment. What tools and strategies would you use to ensure you’re capturing relevant metrics to optimize resource usage and costs?

**Scenario 32**: A company is running high I/O workloads on a storage-optimized instance but is unsure if it’s cost-effective. How would you identify whether the instance type is over-provisioned and recommend a more appropriate instance type or storage option?

**Scenario 33**: You have a set of on-demand instances running 24/7 in a development environment. How would you design a strategy to reduce costs by managing instance lifecycles more effectively?

**Scenario 34**: Describe how you would implement automated scaling policies that account for both cost and performance for a stateless, cloud-native application hosted on AWS.

**Scenario 35**: How would you optimize a complex application that runs a mix of on-demand, reserved, and spot instances across multiple availability zones to reduce costs without compromising on performance?

**Automation and Scripting**

**Scenario 36**: You need to write an automation script to deallocate idle resources in an AWS environment. How would you handle edge cases where certain resources should not be shut down even if underutilized?

**Scenario 37**: Write a script that periodically checks for and deallocates orphaned resources (e.g., EBS volumes, Elastic IPs, etc.). How would you ensure that this script doesn’t accidentally delete resources in use?

**Scenario 38**: Your company wants to optimize the deployment of containers on AWS Fargate to reduce costs. How would you approach tuning task and service configurations to achieve this?

**Scenario 39**: You need to create a CloudFormation template that provisions a cost-optimized multi-tier web application on AWS. What key design principles would you apply to ensure performance isn’t compromised?

**Scenario 40**: You’re tasked with writing an automated tool to adjust resource allocations based on utilization thresholds. How would you balance performance, availability, and cost efficiency in your design?

**Advanced Optimization Techniques**

**Scenario 41**: A machine learning team is using GPU instances to train models, but the team reports under-utilization. How would you approach optimizing instance usage and minimizing costs, considering performance needs?

**Scenario 42**: You’re responsible for a legacy monolithic application that’s being migrated to the cloud. What steps would you take to ensure a smooth migration while minimizing

**Scenario 43:** Your company runs periodic batch processing workloads that require high throughput but are currently over-provisioned in terms of compute resources. How would you redesign the batch processing architecture to use spot instances or serverless solutions to reduce costs without sacrificing throughput?

**Scenario 44 :** You’re optimizing a distributed system running large-scale simulations. The team reports high costs due to unnecessary resource contention and excessive inter-node communication. How would you optimize the system’s architecture to minimize inter-node traffic and reduce operational costs?

**Scenario 45:** A critical AI inference workload uses FPGA instances, but the cost is becoming prohibitive. How would you evaluate whether to switch to other specialized instance types (like GPU or CPU instances) while ensuring the performance required for real-time inferences remains intact?

**Disaster Recovery & High Availability**

**Scenario 46:** Your company has a mission-critical application that must be highly available across multiple regions. You are tasked with designing a disaster recovery (DR) plan. How would you balance the cost of a hot DR solution (active-active) with performance and availability needs?

**Scenario 47:** You’re tasked with setting up a scalable, highly available architecture for an e-commerce platform that must handle unpredictable traffic surges. Describe how you would configure load balancing, auto-scaling, and failover mechanisms while keeping infrastructure costs at a minimum.

**Scenario 48:** In a hybrid cloud setup, your company wants to ensure seamless failover between on-premises and cloud environments in case of disaster, but without maintaining full replicas in both environments due to cost concerns. How would you design a cost-effective disaster recovery strategy in this scenario?

**Scenario 49:** You’re responsible for an application that requires both high availability and low latency, but your budget is constrained. How would you design a multi-region, fault-tolerant architecture using AWS services like Route 53, RDS Multi-AZ, and S3, ensuring minimal failover downtime and cost?

**Scenario 50:** Your organization’s cloud bill is skyrocketing due to redundant high-availability setups across multiple regions. How would you consolidate the infrastructure to reduce redundancy and costs while still maintaining a strong disaster recovery posture?