AWS Certified Solutions Architect: Associate - 2.0 Amazon S3 and Glacier Storage

filename: amazon-acsaa-2-2-amazon_s3_advanced
Title: Amazon S3 Advanced
Subtitle: AWS Certified Solutions Architect: Associate

2.2 Amazon S3 Advanced Features

- Prefixes and Delimiters
    - Used to emulate a file hierarchy
    - Directories/Folders don't actually exist
- Storage Classes
    - S3 Standard
        - High performance, low latency file access
    - S3 Standard-IA
        - S3 Standard Infrequent Access
        - Long lived, but less frequently accessed data
        - Lower per-GB storage per month
        - Minimum object size of 128KB
        - Minimum duration of 30 days
    - S3 Reduced Redundancy Storage (RRS)
        - Lower durability
        - 99.99% vs 99.999999999%
    - Glacier
        - Extremely low-cost
        - Extremely infrequent access
        - Data retrieval takes 3-5 hours
        - Can be deployed as a class in S3, or stand-alone
- Object Lifecycle Management
    - Similar to automated storage tiering
        1. Data stored in S3 Standard for 30 days
        2. Data moved to Standard-IA for 90 days
        3. Data moved to Glacier for 3 years
        4. Data deleted
    - Can be applied to an entire bucket, or a prefix
- Encryption
    - Data in motion secured with SSL API endpoints
    - Data at rest requires configuring encryption
    - Server-Side Encryption (SSE)
        - SSE-S3
            - AWS manages the keys
            - Simply check a box
        - SSE-KMS
            - AWS handles key management
            - You issue the keys
            - Provides additional auditing over SSE-S3
        - SSE-C
            - AWS performs encryption/decryption
            - You manage and provide the keys
    - Client-side Encryption
        - Data is encrypted client-side prior to being uploaded
- Versioning
    - Enabled at the bucket level
    - Every modification is tracked with its own version ID
    - Can be suspended
    - Cannot be disabled
- MFA Delete
    - Multi-factor authentication
    - Once enabled, MFA is required to perform deletes
- Pre-Signed URLs
    - Protection from "content scraping"
    - Leave the bucket private
    - Issue pre-signed URLs for object access
    - Pre-signed URLs are only valid for a defined period of time
    - Must be re-issued when no longer valid
- Multipart Upload
    - Allows better performance when uploading large data
    - Three stages

        - Initiation
        - Uploading
        - Completion
    - Useful for data over 100MB
    - Required for data over 5GB
    - Enabled by default by most clients (e.g. AWS CLI)
- Range GETs
    - Allows retrieving only part of a file
    - Equivalent to multi-part downloads or file resume
- Cross-Region Replication
    - Asynchronous replication between buckets in two different regions
    - Data, metadata and ACLs are all replicated
    - Requires versioning on source and destination
    - Useful when you want data close to the consumer
    - NOTE: Only replicates new objects
        - Existing objects must be manually copied over the first time
- Logging
    - Off by default
    - Logs can be stored in the same or a different bucket
    - Information
        - Requestor account/IP
        - Bucket name
        - Request time
        - Action (GET, PUT, LIST, etc)
        - Response status or error code
- Event Notifications
    - Useful for workflows and alerts
    - For example, can trigger media transcoding
    - Defined at the bucket level
    - Triggers
        - Object created
        - Object deleted
        - RRS object lost
    - Multiple transport options
        - Simple Notification Service (SNS)
        - Simple Queue Service (SQS)
        - AWS Lambda (can trigger a Lambda function directly)
- Best Practices, Patterns and Performance
    - Backup destination for on-premises hybrid-cloud data
    - Blob storage for a database like DynamoDB or RDS
    - Data rates up to 100 requests per second supported by defaut
        - Higher rates are supported
        - However, CloudFront is recommended