

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

filename: amazon-acsa-2-1-amazon_s3_basics

Title: Amazon S3 Basics

Subtitle: AWS Certified Solutions Architect: Associate

2.1 Amazon S3 Basics

- Object Storage vs Block and File Storage
 - Block Storage
 - Data is managed as raw storage on the disk
 - Ordered by fixed-size blocks on the disk
 - File Storage
 - Data is managed at the OS level
 - Ordered by a named hierarchy
 - Object Storage
 - Data is managed via an API over HTTP/HTTPS
 - Data is not tied to a disk or server
 - Theoretically limitless storage
- Buckets
 - Container / Web Folder
 - Form a namespace used to access files
 - Must be unique across all of AWS
 - Best practice is to follow DNS naming conventions
 - Using your domain name in the bucket name helps to achieve uniqueness
- AWS Regions
 - Buckets are created in specific regions
 - Placement is usually based on proximity to the user
- Objects
 - Entities (files) in S3
 - Can store any data
 - Up to 5TB per entity
 - Unlimited per bucket
 - Made up of data and metadata
 - Metadata is made up of name/value pairs
 - System metadata
 - Date modified
 - Size
 - Etc
 - User metadata
 - Custom values
- Keys
 - Unique identifiers that identify stored entities
 - Equivalent of a filename
 - Unique only within a bucket
- Object URL
 - URL used to access the entity
 - `<web_services_endpoint>/<bucket_name>/<object_key>`
 - `https://s3.amazonaws.com/sample123.itpro.tv/heartbeat.php`
- Operations
 - Actions available within the API
 - Create/Delete bucket
 - Write an object
 - Read an object
 - Delete an object
 - List keys in a bucket
- REST Interface
 - Representational State Transfer API
 - HTTP methods used to interact with the S3 API
 - Usually accessed through the SDK, Web UI, AWS CLI or other utilities
- Durability and Availability
 - Durability
 - *Will my data still be there in the future?*
 - Availability
 - *Can I access my data right now?*
 - AWS SLI
 - 99.999999999% Durability
 - If you store 10,000 objects, you could lose 1 every 10,000,000 years

- 99.99% Availability
 - 52 minutes 35.7 seconds per year
 - Reduced Redundancy Storage (RRS)
 - 99.99% Durability
- Data Consistency
 - PUT for a new file
 - uses *read-after-write consistency*
 - PUT for an existing file and DELETE
 - uses *eventual consistency*
 - Subsequent reads could get old data if replication is not complete
 - Reads are consistent
 - You either get new or old data.
 - Never a mix
- Access Control
 - Access restricted to owner by default
 - S3 Access Control Lists
 - Course-grained
 - Legacy mechanism
 - Typically only used for logging and static web hosting
 - Permission
 - READ
 - WRITE
 - FULL-CONTROL
 - S3 Bucket Policies
 - Fine-grained
 - Recommended mechanism
 - Associated with the bucket
 - IAM Policies
 - Similar to bucket policies
 - Associated with an IAM principle instead of the bucket
- Static Web Hosting
 - Support for non-dynamic content
 - Every object has a URL already
 - Use a CNAME to point to the URL
 - Enabling static web hosting changes the URL
 - `<bucket_name>.s3-website-<region>.amazonaws.com`
 - CNAMEs can't point to directories