S3 - Simple Storage Service

AWS S3 - Simple Storage Service what is it? and where would you use it? Also take a look at an architecture using S3.

Fundamentally there are two types of storage:-

1. Object-Based Storage
2. Block Based Storage

Simple Storage Service (S3) - Object-Based Storage

Provides developers and IT Teams with secure durable, highly scalable object storage. Easy to use simple web interface to store and retire any amount of date from anywhere on the web.

It is a place to store your files on the AWS cloud Dropbox was born. By simplifying the User interface of S3. Think about Dropbox as a layer built on top of S3.

Data is spread across multiple devices and facilities

Think about S3 to store your photos or files.

1. Object Base Storage
2. Unlimited Storage
3. Files are Stored in Buckets/Folders
4. Names must be unique globally
5. Every time you have a successful upload you get a http 200 code back

S3 Is primarily used for:-

1. Store and Backup
2. Application File Hosting
3. Media Hosting
4. Software Delivery

4. Software Delivery

5. Storing AMI's and Snapshots

Data consistency Model – S3

Read after write consistency for PUTS of new objects Eventual consistency for overwrite PUTS and deletes (i.e. takes time to propagate)

Objects Consists of the following

1. Key – this is simply the file name of the object

2. Value – the data and it is made up of a sequence of bytes.

3. Versioning – which version of the object is this

4. Meta Data: Data about data, the data about the data file you are storing.

Think, if you are storing a music track/song. This would have metadata like the information of the singer, the year it was released, the name of the album etc.

Sub resources –

1. Access Control list – this determines was can access the file on S3. This can be done at the file level or at the Bucket level.

2. Torrent – supports the Bit torrent protocol

3. Built for 99.99% availability of the S3 Amazon SLA 99.9% platform

4. Durability guarantee – 99.9... (11.9s) (more)?

5. Tiered storage Availability

6. Lifecycle management

7. Versioning

8. Encryption

9. Secure the data using Access control lists and Bucket policies

S3 – IA (Infrequently Accessed) For data that is accessed less frequently but requires rapid access when needed. This here costs lesser than S3 but you are charged for the retrieval of the data.

S3 – RRS (Reduced Redundancy Storage) Basically less durability with the same level of availability.

e.g. This about data you could potentially regenerate like a tax calculation or a

payslip. This is cheaper than the original SS. Suppose you create thumbnails for all your pictures. If you lose a thumbnail you could always regenerate it.

When deciding which storage to use think about the various storage options, their advantages vs disadvantages. Are you optimizing for durability, the frequency of retrieval or availability?

Image comparing the various types of S3

	Standard	Standard - Infrequent Access	Reduced Redundancy Storage
Durability	99.999999999%	99.999999999%	99.99%
Availability	99.99%	99.9%	99.99%
Concurrent facility fault tolerance	2	2	1
SSL Support	Yes	Yes	Yes
First byte latency	Milliseconds	Milliseconds	Milliseconds
Lifecycle Management Policies	Yes	Yes	Yes

Charging Model

1. Storage
2. Number of requests
3. Storage Management Pricing
4. Add metadata to see usage metrics.

Transfer Acceleration - Enables fast, easy and secure transfers of your files over long distances between your end users and an S3 bucket.

Transfer acceleration takes advantages of Amazon cloud front's globally distributed edge locations. As the data arrives at an edge location, the data is routed to Amazon S3 over an optimized network path.

Think of transfer acceleration as a combination of S3 + CDN natively supported by this Service. Basically, every user ends up going through the closest possible edge location which in turn talks to the actual S3 bucket.

Recap - S3

S3 Storage Classes

1. S3 (Durable, immediately available and frequently accessed)
2. S3 – IA (durable, immediately available, infrequently accessed.
3. S3 Reduced Redundancy Storage (Used for data that is easily reproducible, such thumbnails)

Core fundamentals of S3 objects

1. Key: Name of the object These are store in alphabetic order
2. Value: The data itself Version ID: The version of the object
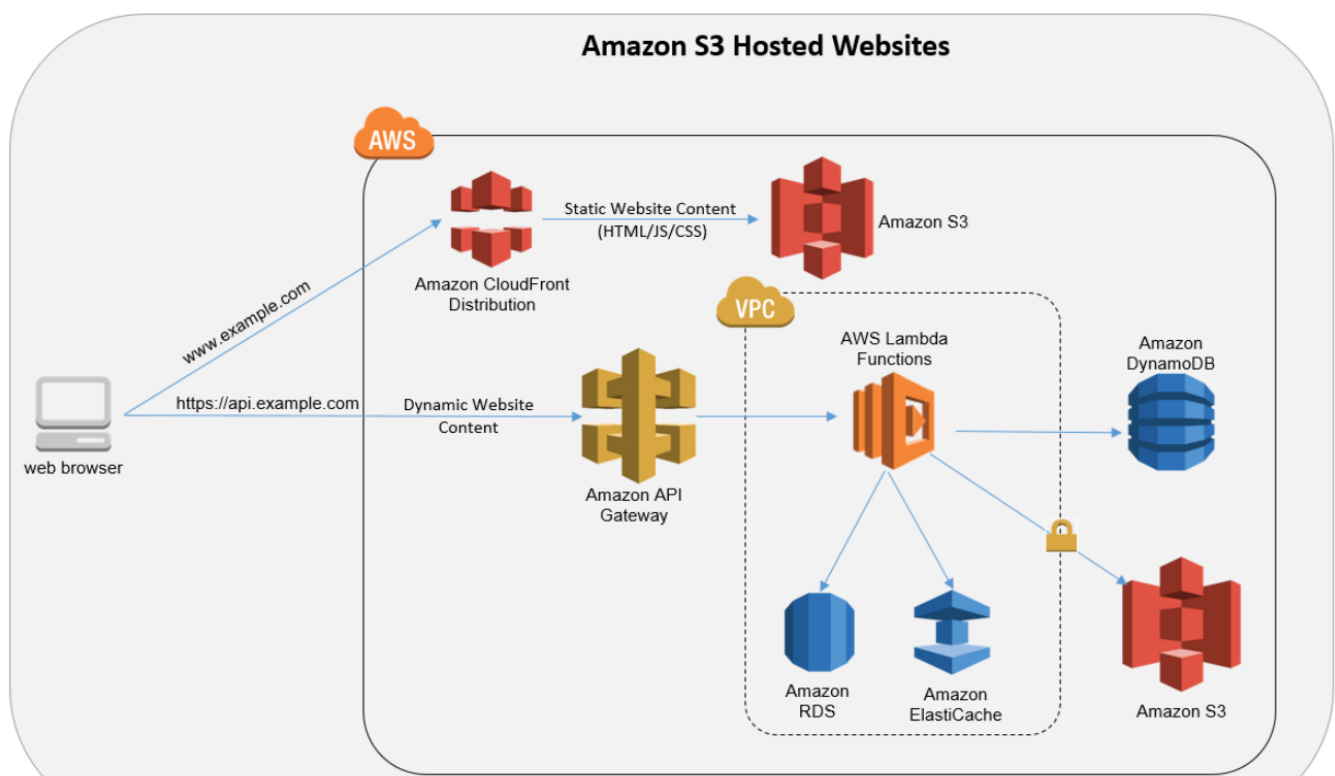3. Meta Data: The various attributes of the data

Sub resources

1. ACL: Access control lists
2. Torrent: bit Torrent protocol

Cross region Replication

This basically means that if you have this turned on then for a bucket AWS will automatically make a bucket available across 2 or more regions.

Example of an Amazon S3 Hosted Website Architecture

Securing your S3 Buckets

1. By default, all buckets are private
2. You can set up access control to your using
3. Bucket Policies
4. Access control lists (ACL)
5. S3 buckets can be configured to create access logs

Encryption for S3

1. In-transit
2. SSL/TLS (using HTTPS)
3. At Rest
4. Server Side Encryption
5. S3 Managed keys – SSE-S3
6. Server side Encryption
7. Key Management Service – Managed Key
8. SST – KMS
9. Client Side Encryption