


AWS Simple Storage Service - S3

TCSS 562 - Software Engineering for Cloud Computing
Spring 2017

Inno Irving V. Estrera
Viet Nguyen
Swetha Chinthapati



3 W's of S3



- What?
 - Amazon S3 provides storage through web services interfaces (REST, SOAP, and BitTorrent).
- Why?
 - To store and retrieve any amount of data, at any time, from anywhere on the web to make web-scale computing easier for developers.
- How?
 - Using and Object Storage Architecture

Object Storage Architecture

- Manages data as objects
- Allows retention of massive amounts of unstructured data.
- Advantages:
 - Ease of use
 - Unlimited scalability
 - Access via Internet protocols rather than storage commands,
 - Custom metadata
 - Best suited as cloud storage

History



CAUTION

History
may lead
to new
insights!!!

2006

- **S3 Introduction**
- Cost is 15 cents per gigabyte per month
- Maximum object size is 5GB

2008

- S3 cost is split into usage tiers
- With larger storage getting as low as 12 cents per GB month.
- Initial tier (0-1TB) remains 15 cents.

2010

- Reduced Redundancy Storage Launched.
- Storage costs to 10 cents per gigabyte per month

History - 2

2010

- IAM Support added.
- Maximum object size increased from 5GB to 5TB
- 0-1 TB -> 14 cents now

2011

- 5GB free S3 storage.
- With larger storage getting as low as 12 cents per GB month.
- Initial tier (0-1TB) remains 15 cents.

2012

- 0-1 TB -> 12.5 cents -> 9.5 cents per month
- Amazon Glacier Support added

History - 3

2014

- 0-1 TB 8.5 cents per month.
- Free usage up to 5GB

2015

- S3 now supports cross-region replication.
- AWS Cloud watch Integration
- New storage class STANDARD_IA (IA, for infrequent access).

2016

- Amazon S3 Analytics – Storage Class Analysis
- Object tagging

Key Features



- Reduced Redundancy Storage
- Bucket Policies
- AWS Identity and Access Management
- Access Control Lists
- Versioning
- Operations

Storage Classes



Standard



Standard –
Infrequent Access



Amazon Glacier

Active data

Infrequently accessed data

Archive data

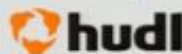
S3 Standard

- Bigdata analysis
- Content distribution
- Static website hosting

The logo for FINRA, featuring the word "FINRA" in blue and a blue triangle pointing to the right.The logo for Netflix, featuring the word "NETFLIX" in white on a red background.

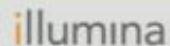
Standard - IA

- Backup & Archive
- Disaster recovery
- File sync & share
- Long-retained data

The logo for Pinterest, featuring the word "Pinterest" in its signature red script font.The logo for Hudl, featuring an orange circular icon with a white shape inside, followed by the word "hudl" in a bold, lowercase sans-serif font.

Amazon Glacier

- Long term archives
- Digital preservation
- Magnetic tape replacement

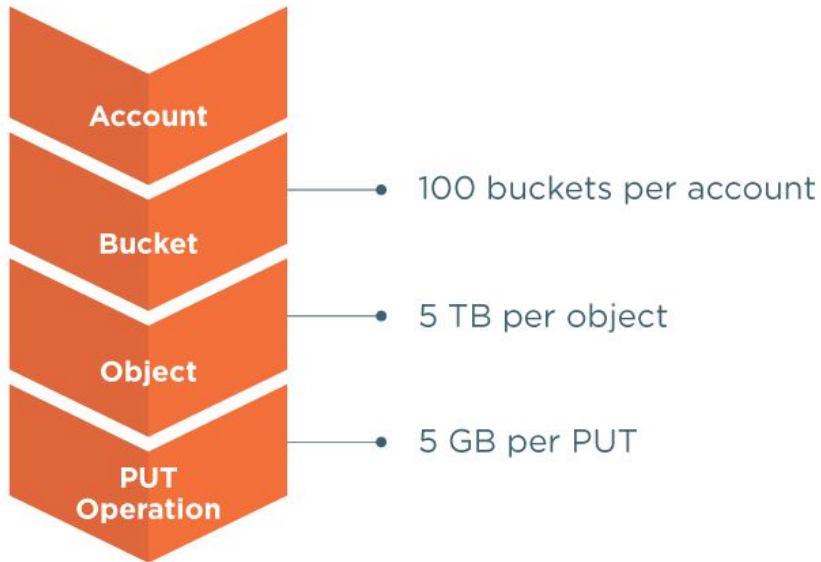
The logo for Illumina, featuring the word "illumina" in a lowercase sans-serif font with a vertical bar to the left of the "i".The logo for Sony DADC New Media Solutions, featuring the text "Sony DADC" in a bold sans-serif font above "NEW MEDIA SOLUTIONS" in a smaller, all-caps sans-serif font.

Use cases for S3

- As a Backup System
- Infrequent access, reduced redundancy and Archival storage
- Host static website
- Create private local Repository
- Host Apache Hadoop
- Improve File Sharing

Advantages

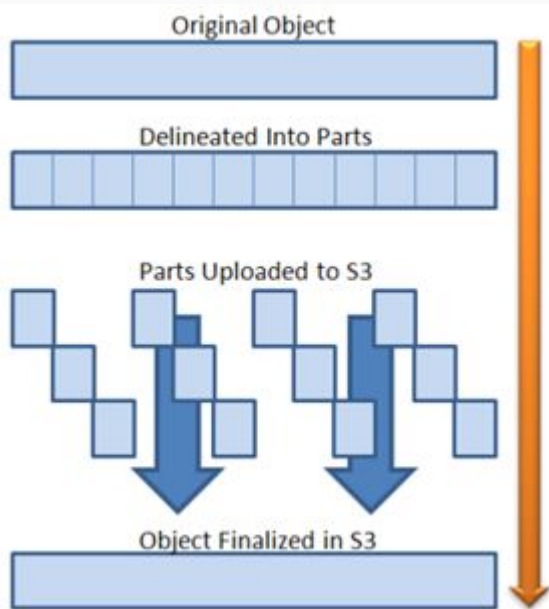
S3 Maximums



Scalable

- Pay per use, control demands. By default, customers can provision up to 100 buckets per AWS account. However, you can increase your Amazon S3 bucket limit by visiting [AWS Service Limits](#).
- Can store as much data as you want and access it when needed. You can stop guessing your future storage needs and scale up and down as required, dramatically increasing business agility
- The total volume of data and number of objects you can store are unlimited. Individual Amazon S3 objects can range in size from a minimum of 0 bytes to a maximum of 5 terabytes. The largest object that can be uploaded in a single PUT is 5 gigabytes.

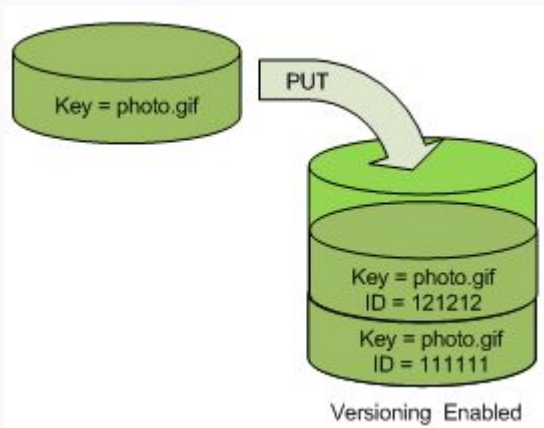
Advantages



Download/Upload

- Download your data any time you like or allow others to do the same. Faster upload and download with Parallelize multi-parts PUTs/GET.
- Increase aggregate throughput by parallelizing PUTs on high-bandwidth networks

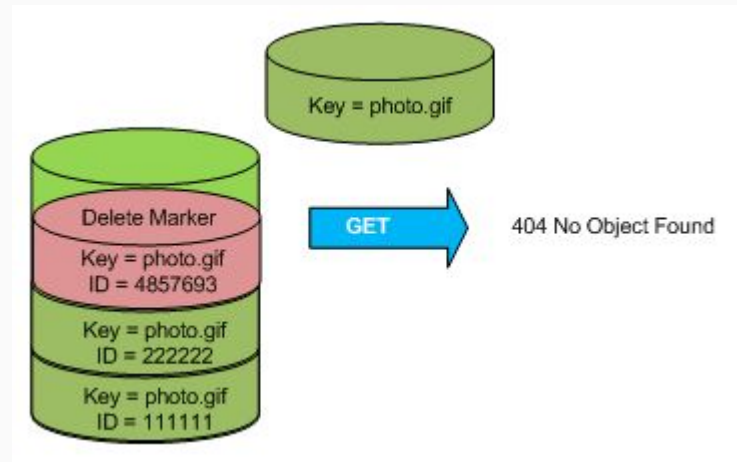
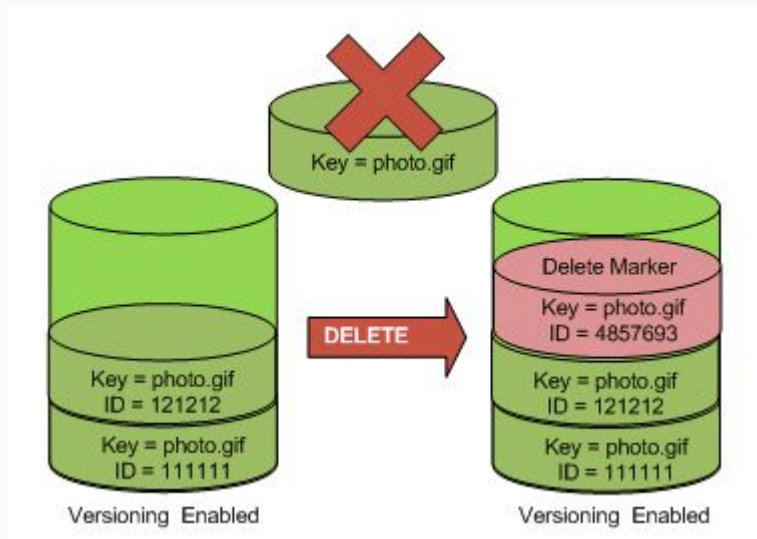
Advantage(cont.)



Versioning

- Enables you to keep multiple versions of an object in one bucket
- This functionality prevents you from accidentally overwriting or deleting objects
- Versioning gives users the opportunity to retrieve a previous version of an object

Advantages (cont.)



Advantages(cont.)

Choice of storage classes on Amazon S3



Standard



Standard –
Infrequent Access



Amazon Glacier

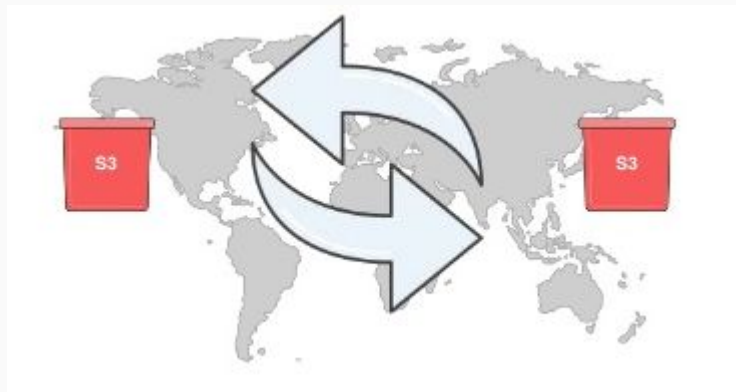
Active data

Infrequently accessed data

Archive data

- Easy to transfer or switching between regular accessed and irregular accessed data
- Taking the advantages of lower cost storage for infrequent data
- Backing up data, long term archive data at very low cost

Advantages(cont.)



S3 cross-region replication

- Automated, fast, and reliable asynchronous replication of data across AWS regions.
- Security- remote replicas managed by separated AWS accounts
- Disaster Recovery: can build rich client-side web applications with Amazon S3 and selectively allow cross-origin access to your Amazon S3 resources.
- Low latency

Disadvantages

- Accessibility: if there is no internet connection then there is no accessing to data
- Data Security: Since it is cloud base, there are concerns with safety and privacy of important data stored remotely. There are a vulnerabilities that data might be stolen and exposed over the internet.
- Software: To manage and manipulate your files locally, there may need to download the services for every devices .
- Platform dependencies: might not easy to migrate from one to the other systems. Migration may put data in security and privacy risk

Disadvantages

- Bucket ownership is not transferable; however, if a bucket is empty, you can delete it. After a bucket is deleted, the name becomes available to reuse, but the name might not be available for you to reuse for various reasons.
- Cannot create a bucket within another bucket.
- If it is down, you are down
- Other problems is that **bandwidth cost is very high**; if you are having very large size public keys in S3 and someone gets holds of url of those keys; he may make you bankrupt by continuously fetching data. So **you should be very careful about access specifier and key distribution. Whenever possible, hide your keys**. Also you have to trust other with your critical data; it's outside your firewall.

Integrating S3



Easy to Use

Amazon S3 is accessed simply through the S3 Console, SDKs, or ISV integration. S3 is supported by the AWS SDKs for Java, PHP, .NET, Python, Node.js, Ruby, and the AWS Mobile SDK. The SDK libraries wrap the underlying REST API, simplifying your programming tasks.



Storage Pricing (varies by region)

Region:

	Standard Storage	Standard - Infrequent Access Storage †	Glacier Storage
First 50 TB / month	\$0.023 per GB	\$0.0125 per GB	\$0.004 per GB
Next 450 TB / month	\$0.022 per GB	\$0.0125 per GB	\$0.004 per GB
Over 500 TB / month	\$0.021 per GB	\$0.0125 per GB	\$0.004 per GB

Ex. 100GB x \$0.023/GB = \$2.3 / month for Standard, \$1.25 for Infrequent Access, and \$0.40 for Glacier Storage

Request Pricing

For Requests Not Otherwise Specified Below

PUT, COPY, POST, or LIST Requests	\$0.005 per 1,000 requests
GET and all other Requests	\$0.004 per 10,000 requests
Delete Requests	Free †

For Standard – Infrequent Access Requests

PUT, COPY, or POST Requests	\$0.01 per 1,000 requests
GET and all other Requests	\$0.01 per 10,000 requests
Lifecycle Transition Requests into Standard – Infrequent Access	\$0.01 per 1,000 requests
Data Retrievals	\$0.01 per GB

For Glacier Requests

Lifecycle Transition Requests into Glacier	\$0.05 per 1,000 requests
--	---------------------------

Data Transfer Pricing

Data Transfer IN To Amazon S3

All data transfer in	\$0.000 per GB
----------------------	----------------

Data Transfer OUT From Amazon S3 To

Amazon EC2 in the same region	\$0.000 per GB
Another AWS Region	\$0.020 per GB
Amazon CloudFront	\$0.000 per GB

Data Transfer OUT From Amazon S3 To Internet

First 1 GB / month	\$0.000 per GB
Up to 10 TB / month	\$0.090 per GB
Next 40 TB / month	\$0.085 per GB
Next 100 TB / month	\$0.070 per GB
Next 350 TB / month	\$0.050 per GB

Conclusion

- AWS S3 is a great entry point for new entrant in the industry with minimum possible cost and no worries about maintenance and infrastructure.
- You have access to highly durable storage, high-performance, management tools, and more. All this is available without up-front cost, and you pay for only what you use. These services help organizations move faster, lower IT costs, and scale.
- AWS is designed to work with most of the popular modern systems to power a wide variety of workloads, including web and mobile applications, game development, data processing and warehousing, storage, archive, and many others.

Demo - S3

Questions

