aws

SIMPLE STORAGE SERVICE

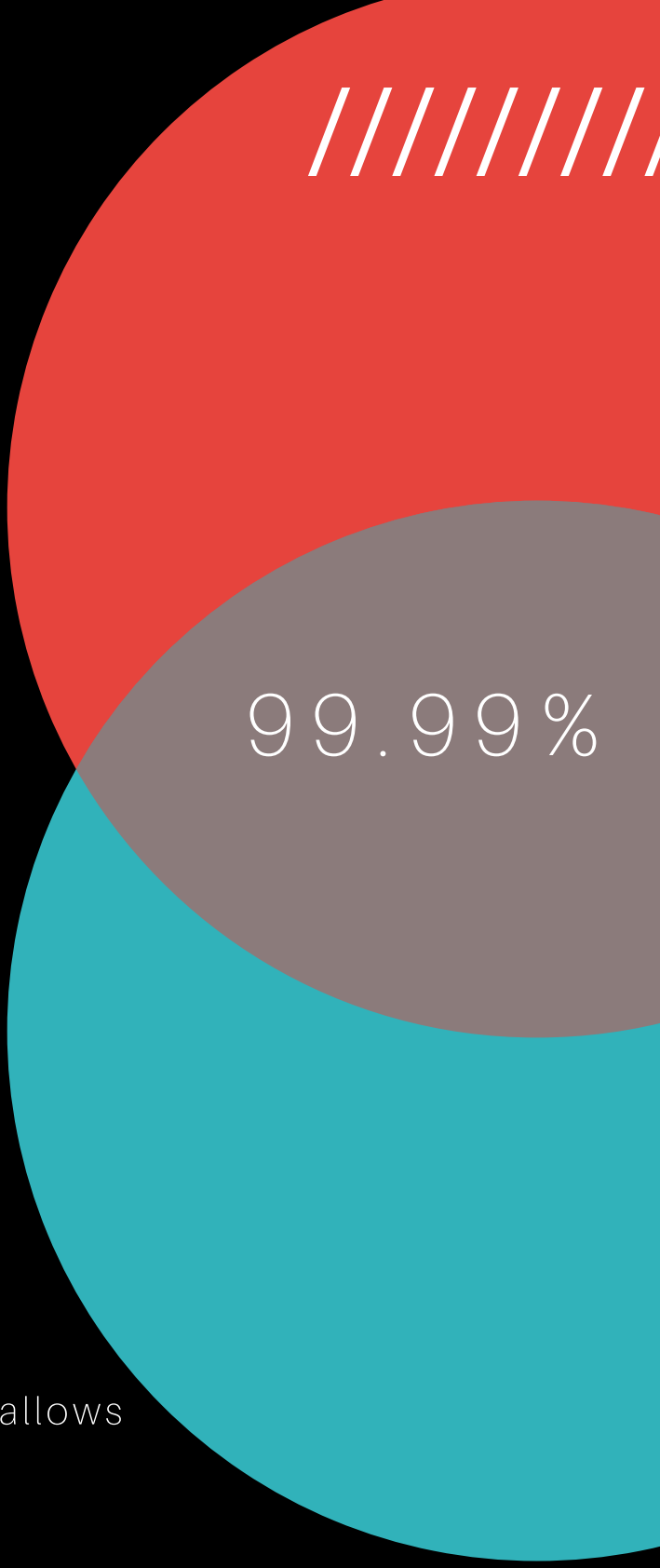
RIYAZ UL HAQUE

·CONSULTADD·

TABLE OF CONTENTS

S3-BASICS

- Built for 99.99% availability for the S3 platform.
- Amazon gurantee 99.999999999 durability for S3 information. (Remember 11 x 9s)
- Tiered storage available.
- Life-cycle management.
- Versioning.
- Encryption.
- Secure your data using Access Control Lists and Bucket Policies



- Remember S3 is object storage i.e. allows you to upload files.
- Files can be from 0 Bytes to 5TB.
- There is Unlimited storage.
- Files are stored in Buckets.
- S3 is universal namespace.
- http://s3-regionname.amazon.com/bucketname

AMAZON-S3
SIMPLE STORAGE SERVICE

Object storage built to store and retrieve any amount of data from anywhere

amazon | S3

Amazon S3 is cloud storage that provides developers and IT teams with secure, durable, highly scalable object storage. With Amazon S3 you can store and retrieve any amount of data from anywhere on the web using simple web services interface- websites and mobile apps, corporate applications, and data from IoT sensors or devices at a massive scale.

- Amazon S3 gives customers flexibility in the way they manage data for cost optimization, access control, and compliance.
- Amazon S3 works with your business processes and even enhances those processes by allowing anyone to securely run queries on his or her data without moving it to a separate analytics platform.

A bucket is a logical unit of storage in Amazon Web Services (AWS) object storage service, Simple Storage Solution S3. Buckets are used to store objects, which consist of data and metadata that describes the data.

GETTING STARTED WITH AMAZON S3

Getting started with Amazon S3 is simple.

- Sign Up for the Amazon S3 console
- Create your bucket
- Add an object
- View and Move an object
- Delete an object and Bucket

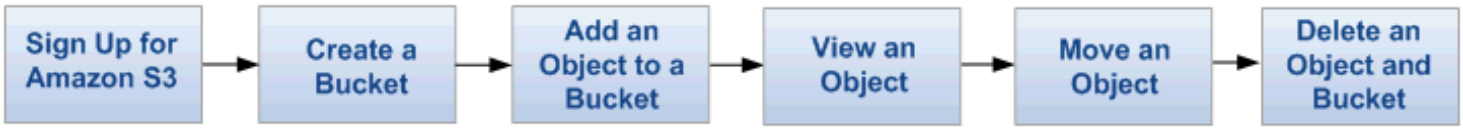


fig: 01

An S3 customer must create a bucket before he or she can store data in Amazon's public cloud and specify access privileges for the bucket by using the AWS Policy Generator. Although customers are not charged for creating buckets, they are charged for storing objects in a bucket and for transferring objects in and out of buckets.

There is no limit to the number of objects a customer can store in a bucket, but each AWS account can only have 100 buckets at one time.

BUCKET RESTRICTIONS AND LIMITATIONS

A bucket is owned by the AWS account that created it with following restrictions and limitations.

- You can create up to 100 buckets in each of your AWS accounts.
- Bucket ownership is not transferable.
- There is no limit to the number of objects that can be stored in a bucket.
- No difference in performance whether you use many buckets or just a few.
- Once you have created a bucket, you can't change its Region.
- You cannot create a bucket within another bucket.

RULES FOR BUCKET NAMING

Rules for naming the bucket are very important because once you create an S3 bucket, you can't change the bucket name, so choose the name wisely.

- Bucket names must be unique.
- Bucket names must comply with DNS naming conventions(According to the rules of the new convention).
- Bucket names must be at least 3 and no more than 63 characters long.
- Bucket names must not contain uppercase characters or underscores.
- Bucket names must start with a lowercase letter or number.
- Bucket names cannot contain dashes next to periods (e.g., my-.bucket.com and my.-bucket are invalid)
- Bucket names must not be formatted as an IP address (for example, 192.168.5.4).

Valid Examples

- my-first-bucket-3
- my-project-bucket
- 4my-group

fig: 02

OBJECT STORAGE

Object storage or Object-based storage is a general term that refers to the way in which we organize and work with units of storage, called objects. Every object contains three things.

- Data
- Metadata
- Globally Unique Identifier

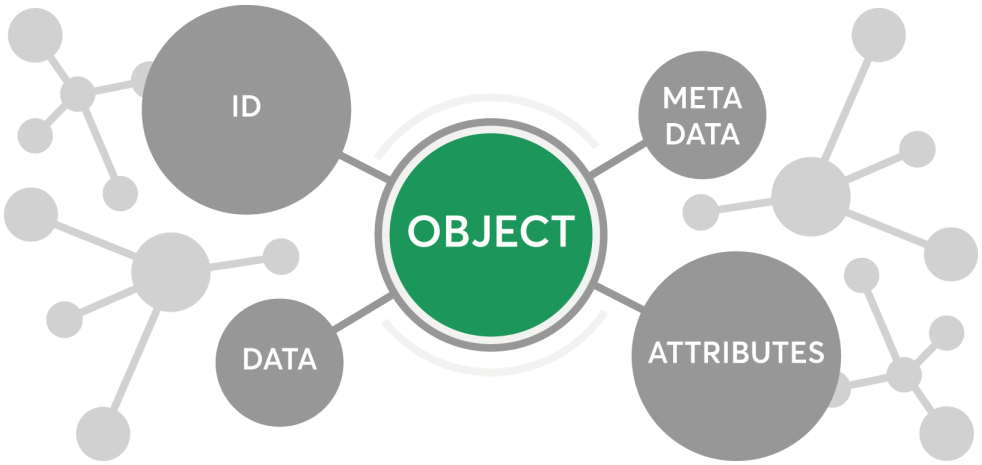


fig: 03

Data

The data can be anything you want to store, from a media file such as a photo to a 400,000-page manual for assembling an aircraft.

Metadata

The metadata is defined by whoever creates the object storage; it contains:

- Contextual information about what the data is
- What it should be used for
- It's confidentiality
- Anything else that is relevant to the way in which the data is used.

Globally Unique Identifier

The identifier is an address given to the object in order for the object to be found over a distributed system. This way, it's possible to find the data without having to know the physical location of the data.

AMAZON S3

RACKSPACE CLOUD FILES

GOOGLE CLOUD STORAGE

AZURE BLOB STORAGE

BLOCK STORAGE

Block storage devices provide fixed-sized raw storage capacity. Each storage volume can be treated as an independent disk drive and controlled by an external server operating system.
Block storage is ideal for databases, since a DB requires consistent I/O performance and low-latency connectivity.

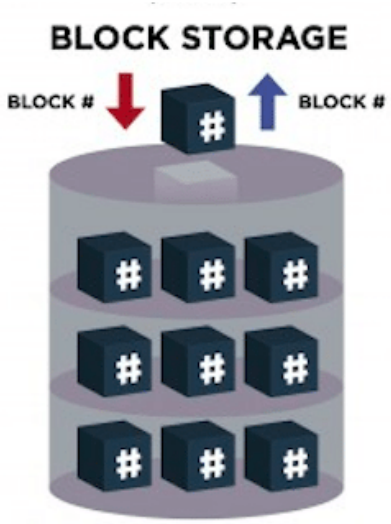


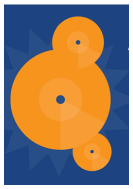
fig: 04

Block storage devices typically are formatted with a file system like FAT32, NTFS, EXT3, and EXT4.

AMAZON EBS

RACKSPACE CLOUD BLOCK STORAGE

GOOGLE PERSISTENT DISKS



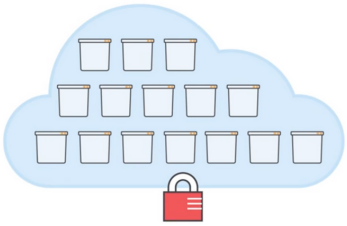
AZURE PREMIUM STORAGE

Block level storage devices are accessible as volumes and accessed directly by the operating system, they can perform well for a variety of use cases. Good examples for block storage use cases are structured database storage, random read/write loads, and virtual machine file system (VMFS) volumes.

BENEFITS

If you're looking for an affordable, multi-purpose means of storing data in the cloud, then Amazon's S3 Simple Storage Service is an option worth considering because of following listed benefits:

Unmatched durability, availability, & scalability
Most comprehensive security & compliance capabilities
Query in place
Flexible management
Most supported by partners, vendors, & aws services
Easy, flexible data transfer



COMPREHENSIVE SECURITY

DIFFERENT AVAILABILITY ZONES

VERSIONING CONTROL

The benefits of Amazon Web Services are directly proportional to the number of ways to use it.

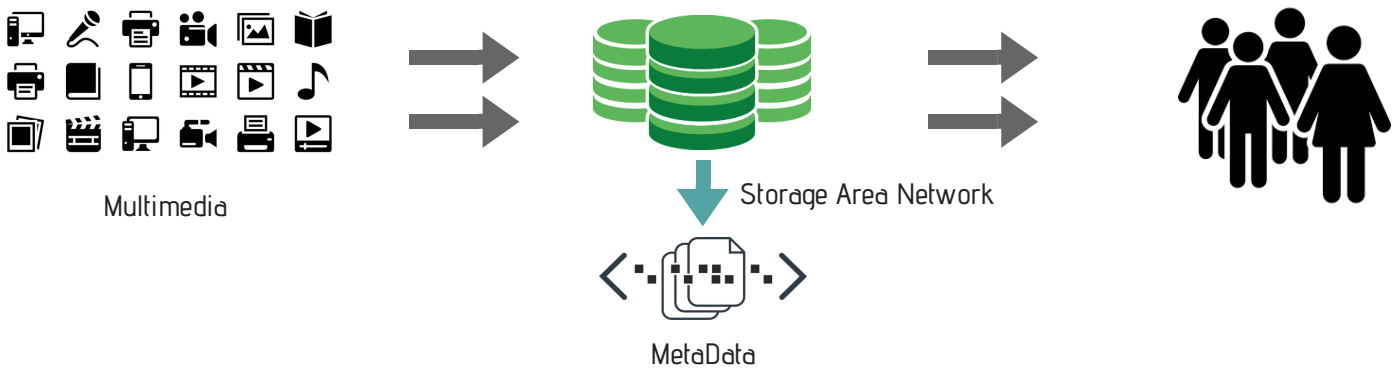
USE CASES

Amazon S3 use case list that will help you take the most out of it in your workflow.

- Backup & recovery
- Data archiving
- Data lakes & big data analytics
- Hybrid cloud storage
- Cloud-native application data
- Disaster recovery

SCENARIO- A Media Company with Large Multimedia Files and Related Metadata

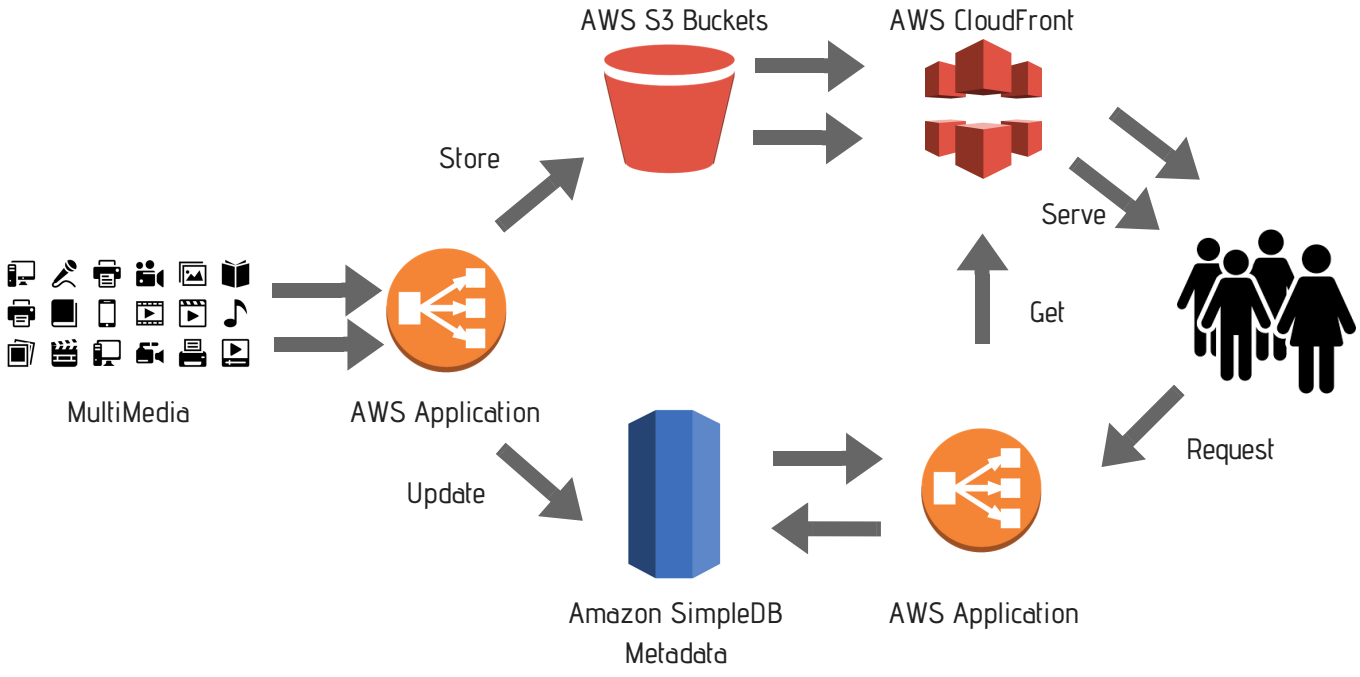
This company provides a tremendous number of multimedia files (audio files, videos, and images), which are stored on internally hosted servers and made available over the Internet. In addition to the multimedia files themselves, each file requires a significant amount of metadata (such as title, author, keywords, size, and so on). The figure below shows the current configuration.



On the plus side, their website is a hit. Both traffic and the number of managed multimedia files are growing rapidly, with tens of thousands of new multimedia files uploaded daily. Unfortunately, internal servers are failing to keep up with this demand—the sheer amount of data will soon overwhelm available disk storage, and the amount of necessary metadata indexing is outstripping processing capacity.

In light of these facts, their architects decide to publish all content to the AWS cloud-based storage, thereby eliminating the need to purchase and maintain internal servers. Instead, users will directly fetch content from AWS-based storage. The combined solution uses Amazon S3, CloudFront, and SimpleDB:

- The application will be built with either the AWS Toolkit for Java or templates for Visual Studio .NET.
- The multimedia content will be stored in Amazon S3.
- After a multimedia artifact has been uploaded, a series of key-value pair entries are created in SimpleDB. These entries will contain metadata related to the multimedia content, and will be created using either the AWS supplied SOAP or REST-based API, or by using the AWS Toolkit for Java or AWS templates for Microsoft .NET.
- When a user wants to retrieve a particular object, the application will first search SimpleDB using the metadata supplied in a lookup form on the website. Access to SimpleDB will be either via direct AWS SOAP or REST API invocations or the development platform plug-ins.
- When a reference to the multimedia entry has been located, a call to Amazon S3 or a corresponding CloudFront distribution will be made to retrieve the multimedia object itself.
- To maximize efficiency when accessing this information, frequently-requested "hot content" will be delivered using Amazon CloudFront via a global network of edge locations. Requests for multimedia objects will be automatically routed to the nearest edge location, thereby making this information available with lower latency and higher data transfer speeds.



CREATE YOUR FIRST BUCKET

Create your first Cloud Storage bucket using Amazon Simple Storage Service (Amazon S3).

STEP 01

Logging in to the Amazon Web Services Console

Sign in to the preview version of the AWS Management Console.

Step 2

Under Storage & Content Delivery, choose S3 to open the Amazon S3 console.

From the Amazon S3 console dashboard, choose Create Bucket.

Step 3

In Create a Bucket, type a bucket name in Bucket Name.

The bucket name you choose must be globally unique across all existing bucket names.

Step 3

Uploading a File to S3

Upload content to your S3 bucket

Step 4

Granting Public Access to an S3 Object

Allow public access to an object (file) in S3

Step 5

Changing Metadata of an S3 Object

Change system and user metadata of an S3 object

Step 6

Deleting an S3 Bucket

Delete an S3 bucket and its content

CASE STUDY

Amazon S3 is the central data hub for Netflix's big data ecosystem

NETFLIX

Netflix is one of the world's largest online media streaming providers delivering almost 7 billion hours of videos to nearly 50 million customers in 60 countries per quarter.

Netflix delivers billions of hours of content to customers around the world from Amazon S3. S3 also serves as the data lake for their Big Data analytics solution.
With Amazon Web Service Netflix quickly deploy thousands of servers and terabytes of storage within minutes.
Users across the globe can stream Netflix shows and movies on the web, on tablets, or on mobile devices such as iPhones.

Netflix currently has over 1.5 billion objects and 60+ PB of data stored in S3.
Netflix ingests, transform, transport, and visualize data, data naturally weaving in and out of Amazon S3. Netflix uses Amazon S3 provides the flexibility to use an interoperable set of big data processing tools like Spark, Presto, Hive, and Pig.
It serves as the hub for transporting data to additional data stores/engines like Teradata, Redshift, and Druid, as well as exporting data to reporting tools like Microstrategy and Tableau.
Netflix now has their own ecosystem of services and tools to manage the data on S3.

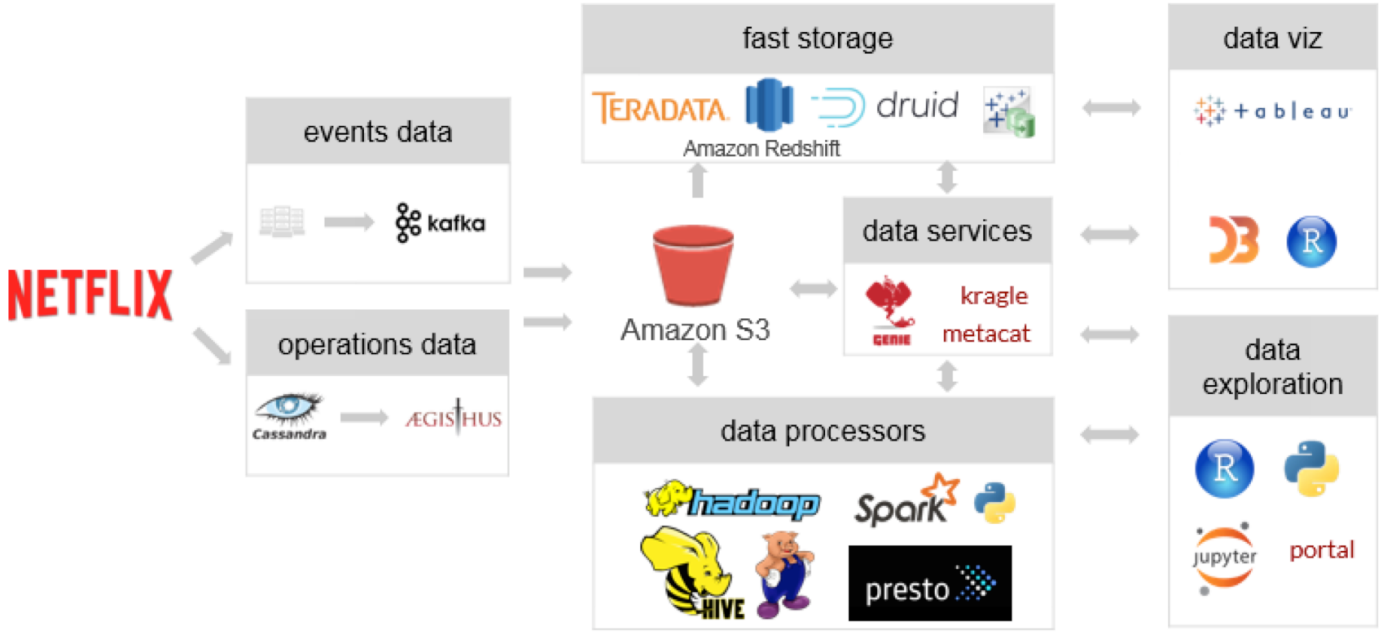


fig: 05

RIYAZ UL HAQUE