



# AWS Foundation

Storage – Simple Storage Service (S3)



# Agenda

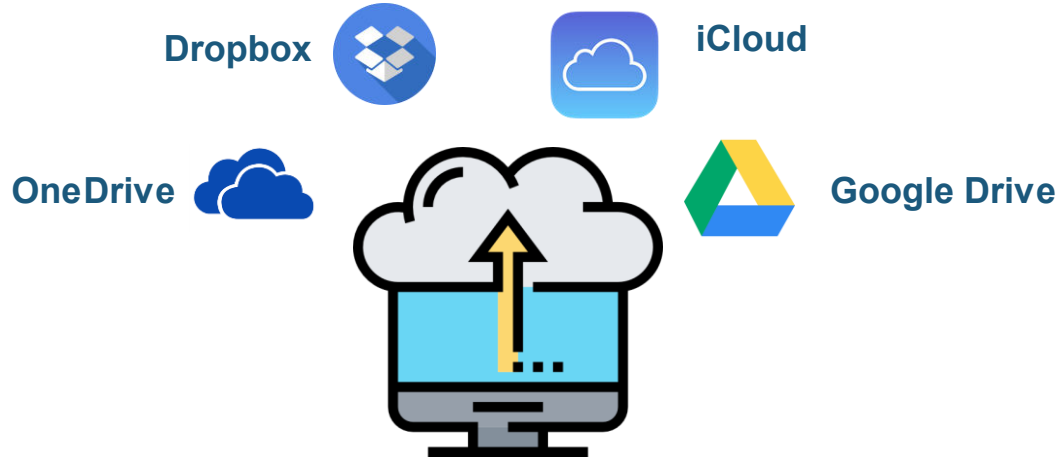


1	Pre-S3: Online Cloud Storage	9	Demo	17	Cross-Region Replication
2	Pre-S3: API	10	Objects	18	Data Encryption
3	S3 Introduction	11	Metadata & Storage Class	19	Server Access Logging
4	S3 Consistency Models	12	Versioning	20	Demo
5	Examples	13	Demo	21	Connect using VPC endpoint
6	Demo	14	Lifecycle Management	22	S3 Pricing
7	Storage Hierarchy	15	Storage Class Analysis	23	Quiz
8	Buckets	16	Demo		

# Pre-S3: Online Cloud Storage

# Online Cloud Storage

Upload files, folders, images, songs, videos from a machine and access it from anywhere in the world.





IntelliPaat



# Pre-S3: API



IntelliPaat

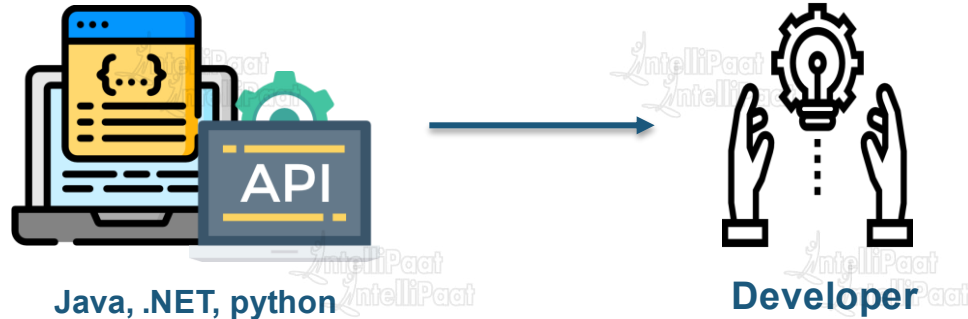


# Application Programming Interface (API)



## What is API?

- ★ An **API** is a list of specifications which describes how information is exchanged between programs.
- ★ Software that wants to access another will call the API published by the other program.



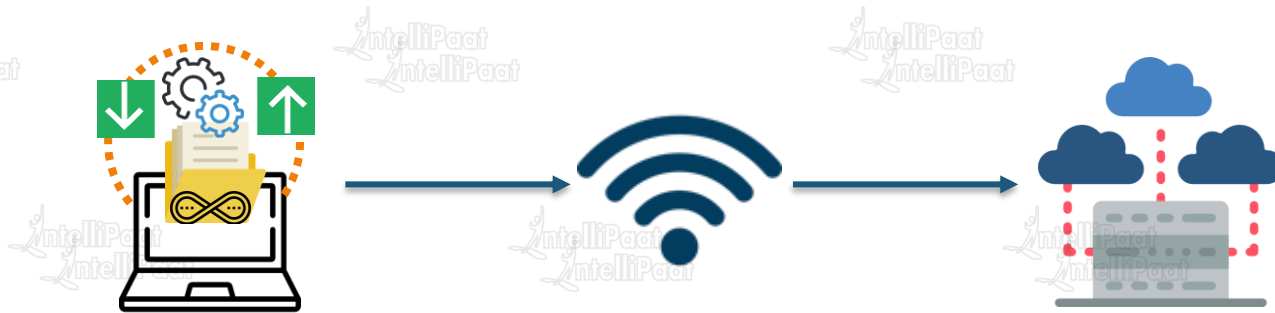
# S3 Introduction

# S3 Concepts



## Simple Storage Service

- ★ Amazon Simple Service (S3) is a storage that can be maintained and accessed over the Internet.
- ★ S3 provides web service which can be used to store and retrieve unlimited amount of data. Same can be done programmatically using Amazon provided APIs.





# S3 Consistency Models

# S3 Data Consistency Model

- ★ S3 provides highly durable and available solution by replicating all data in multiple data centers in a region.
- ★ Data uploaded in a particular region never leaves it.
- ★ Read-after-write consistency.
- ★ Eventual consistency.



# S3 Data Consistency Model

Consistent Read	Eventual Consistent Read
No stale reads	Stale reads are POSSIBLE
Higher comparative read latency	Lower comparative read latency
Read throughput is comparatively lower	Read throughput is highest



**Replication of  
Data in multiple  
Data Centers**



**Data uploaded  
never leaves the  
data center**



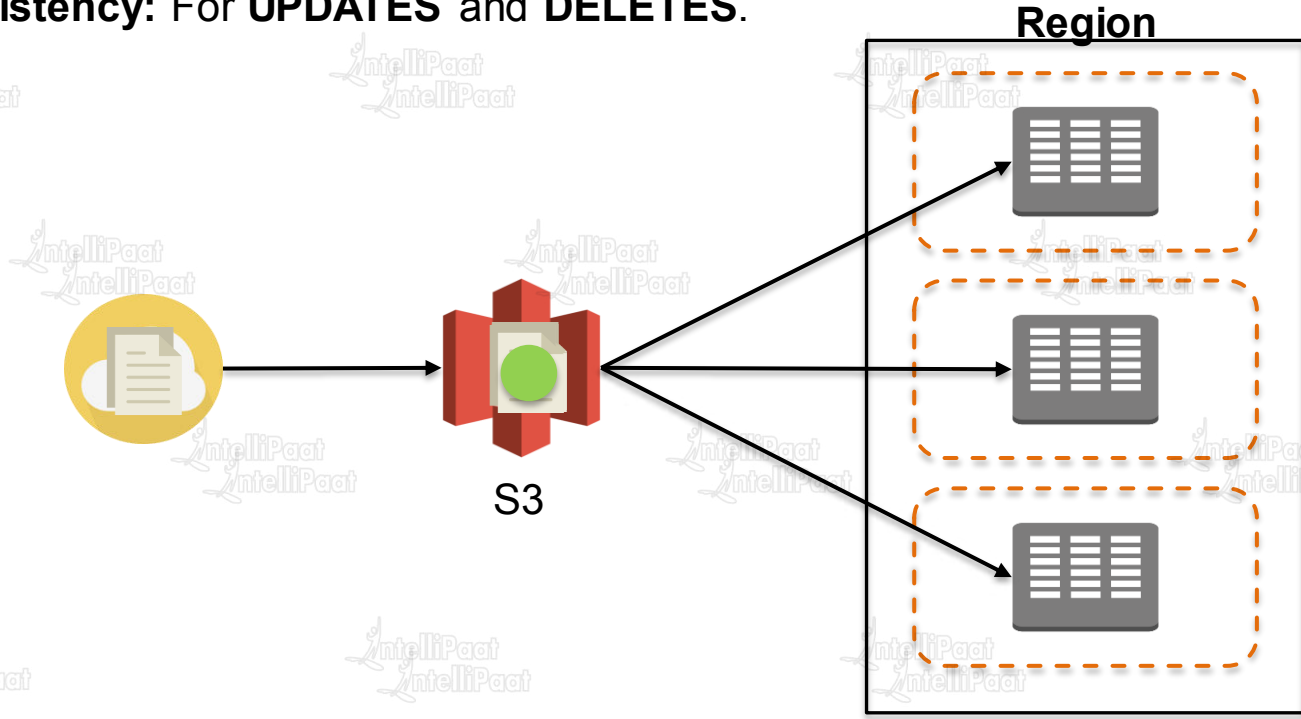
**Read after write  
consistency**



**Eventual consistency**

# Consistency Models

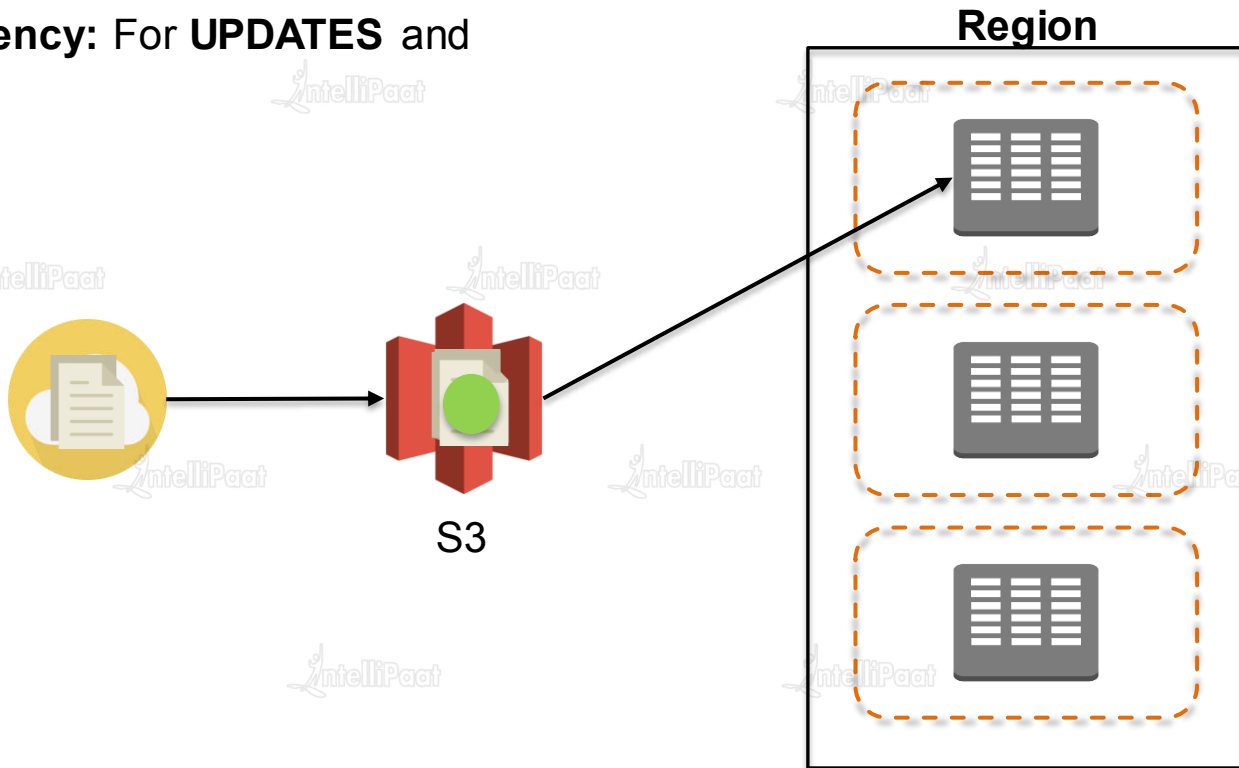
**EVENTUAL consistency:** For **UPDATES** and **DELETES**.



# Consistency Models

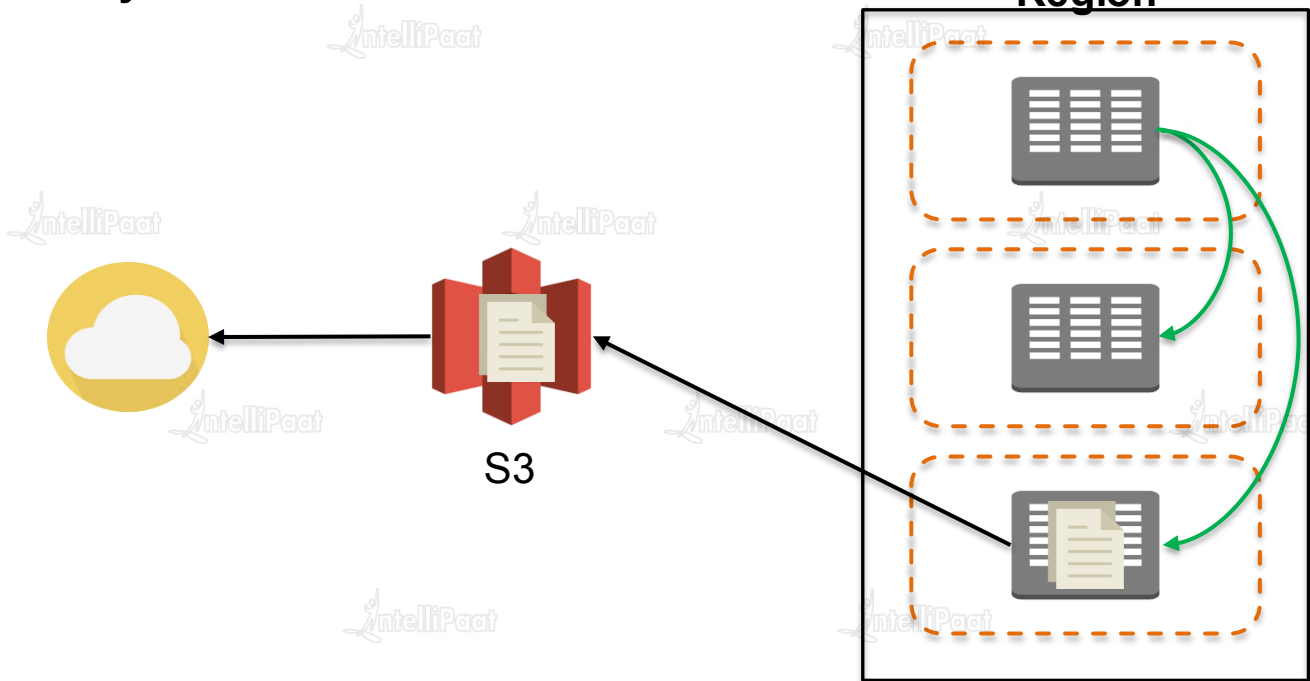


**EVENTUAL consistency:** For **UPDATES** and **DELETES**.



# Consistency Models

**EVENTUAL consistency:** For **UPDATES** and **DELETES**.



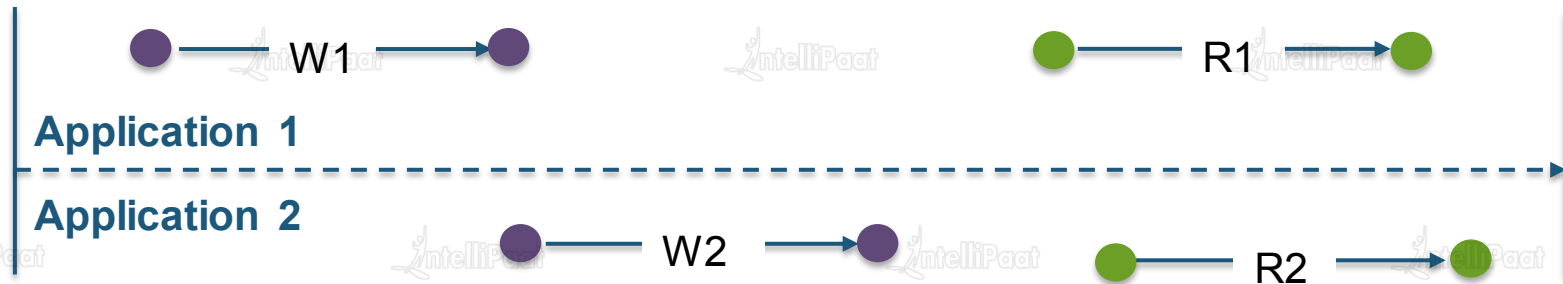
# Examples

# Consistency Examples



## Example 1

### ★ Concurrent applications



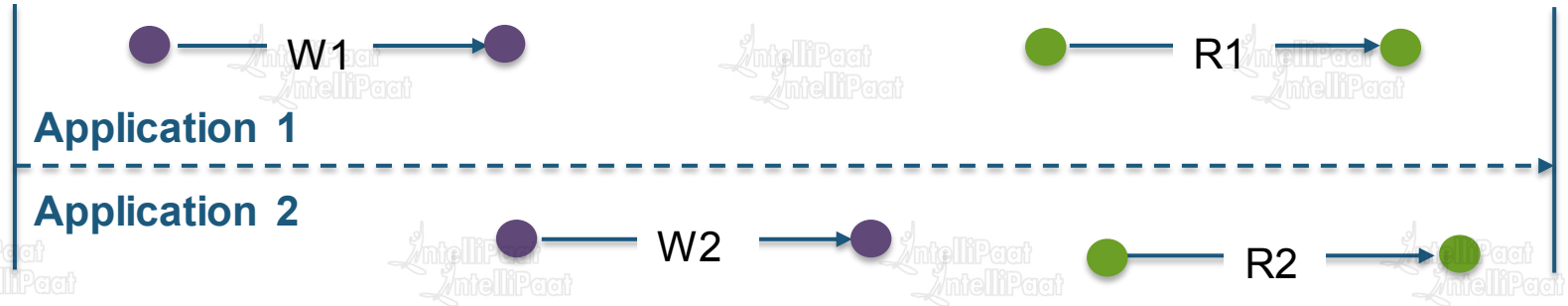
- ★ W1 → Name: “EC2”, W2 → Name: “EBS”
- ★ R1 → consistent – Name: “EBS”, eventual – Name: “EBS” or “EC2” or Nothing.
- ★ R2 → consistent – Name: “EBS”, eventual – Name: “EBS” or “EC2” or Nothing.



# Consistency Examples

## Example 2

### ★ Concurrent applications



- ★ W1 → Name: “EC2”, W2 → Name: “EBS”
- ★ R1 → consistent – Name: “EBS” or “EC2”, eventual – Name: “EBS” or “EC2” or Nothing.
- ★ R2 → consistent – Name: “EBS”, eventual – Name: “EBS” or “EC2” or Nothing.

# Demo

# Demo 1: Upload Files in S3



- ★ Let's upload a few image and text files in S3.
- ★ How to access those files.



IntelliPaat



# Storage Hierarchy



IntelliPaat



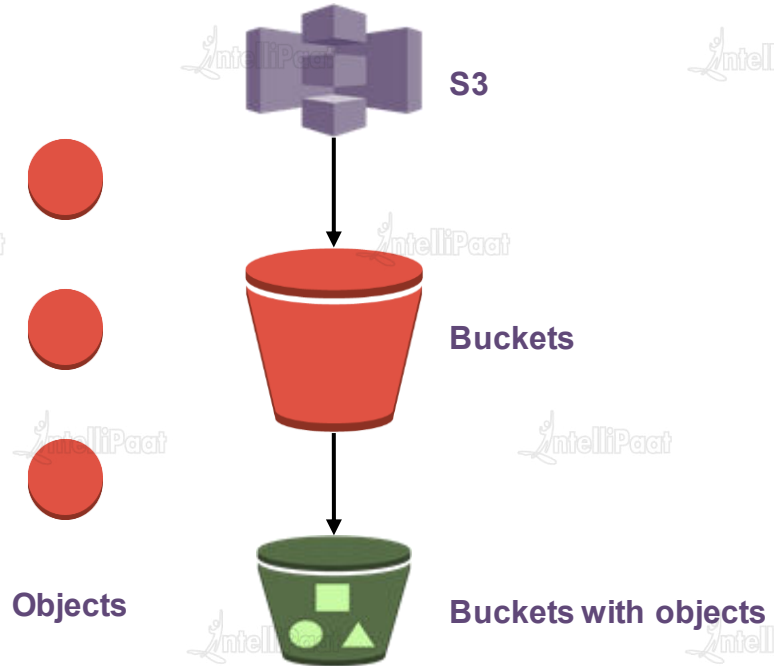
# Storage Hierarchy



S3 follows a storage hierarchy in keeping data (documents, images, videos, files etc.).



Management console or S3 APIs can be used to manage buckets and objects.





IntelliPaat



# Buckets



IntelliPaat



## Bucket count & restrictions

## Communicating using SDK

## Accessing Buckets

## Naming Convention

By default, the maximum number of buckets that can be created per account is 100. For additional buckets, one can submit a service limit increase.



## Bucket count & restrictions

## Communicating using SDK

## Accessing Buckets

## Naming Convention

While using AWS SDKs, the client is created and then this client is used to send request to create bucket. The client is created by specifying an AWS region and then client uses endpoint to communicate with Amazon S3 .

### For Example:

If client is created by specifying the N. Virginia (Default region), then following endpoint is used to communicate Amazon S3:

s3.amazonaws.com

For any other Region:

– s3<region>.amazonaws.com



Bucket count & restrictions

Communicating using SDK

Accessing Buckets

Naming Convention

## Types of URL to access buckets

- ★ Virtual hosted style:  
<http://bucket.s3.amazonaws.com/object> OR  
<http://bucket.s3-aws-region.amazonaws.com/object>
- ★ Path style: <http://s3.amazonaws.com/bucket/object> OR  
<http://s3-aws-region.amazonaws.com/bucket/object>



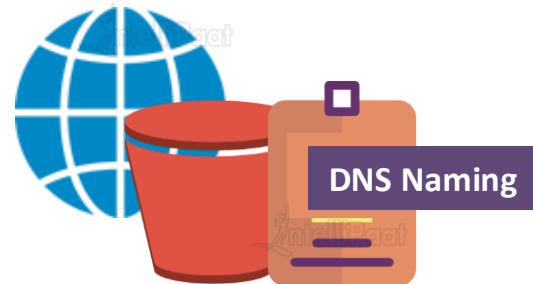
## Bucket count & restrictions

## Communicating using SDK

## Accessing Buckets

## Naming Convention

Bucket names have to be Globally unique irrespective of which region they are created in. As bucket can be accessed using URLs, it is recommended that bucket names follow DNS naming conventions, all letters should be in small.





IntelliPaat



# Objects



IntelliPaat



# Objects



## ★ When there is no folder and object resides in the bucket



<http://my-s3-bucket.s3.amazonaws.com/myobject>

## ★ When there is a folder on console, folder name is used as prefix with object key.



<http://my-s3-bucket.s3.amazonaws.com/myfolder/myobject>

★ Objects are videos, images, documents etc. which are stored inside of buckets.

★ While creating a bucket a name is given, “name” is the object key.

★ There cannot be any sub-bucket or sub-folder inside of a bucket (physically, however folders can be created on the console which provides a logical hierarchy only, which are used as prefixes in the object key.)



IntelliPaat



# Metadata & Storage Class



IntelliPaat



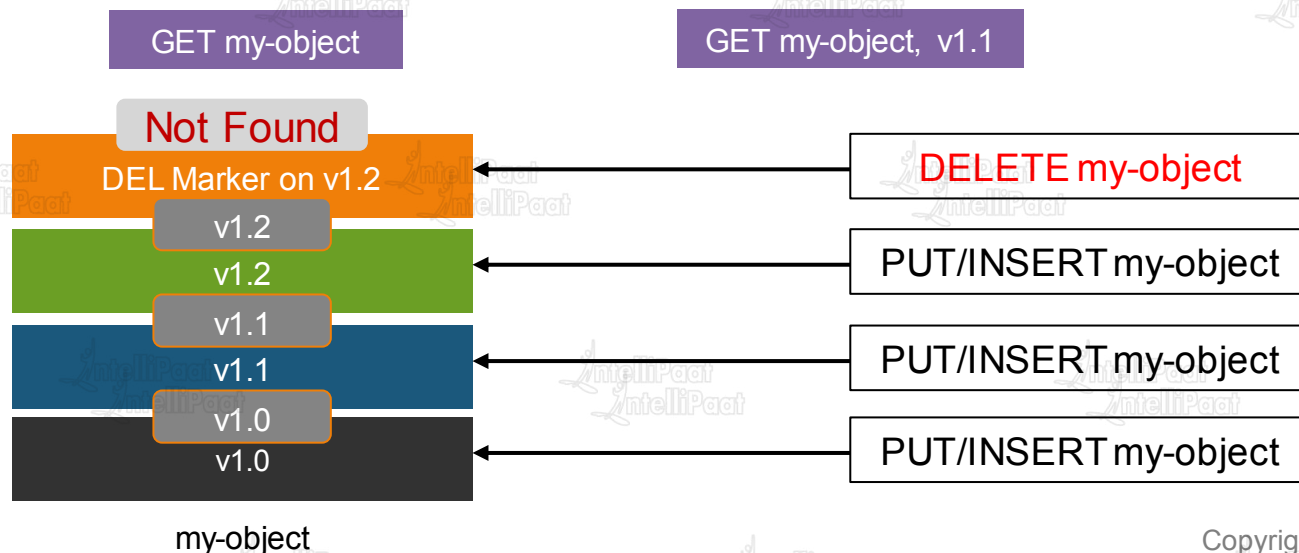
- ★ Object metadata – for each object S3 maintains a set of system metadata.
  - Date: Current date and time.
  - Content-length: Object size in bytes.
  - Last-modified: Object creation or last modified date.
  - x-amz-server-side-encryption: whether encryption is enabled or not.
  - x-amz-version-id: object version.
  - x-amz-delete-marker: Whether the object is a delete marker in case of versioning.
  - x-storage-class: Storage class associated with the object.
- ★ Storage Class – each object has a storage class associated with it.
  - STANDARD: For frequently accessed data. 11 9s of durability, 4 9s of availability.
  - STANDARD IA: For less frequently real-time accessed data. 11 9s of durability, 3 9s of availability.
  - REDUCED REDUNDANCY: For non-critical, reproducible data with lower levels of redundancy than Standard storage class. 4 9s of durability and 4 9s of availability.

# Versioning

# Versioning



- ★ Versioning enables to keep multiple versions of the same object in one bucket.
- ★ Versioning has to be enabled explicitly. Each object has a version ID.
- ★ Existing objects are not overwritten.







IntelliPaat



# Demo



IntelliPaat



Copyright IntelliPaat, All rights reserved



# Demo 3: Objects



- Upload a few objects in the bucket “**aws-foundation-bucket**” created earlier.
- Set permission to everyone for both buckets and objects.
- Check storage class for the object created.
- Enable versioning in the bucket.
  - Add a “**application:AWS-3**” and “**Content-type**” metadata in one of the objects.



IntelliPaat



# Lifecycle Management



IntelliPaat



- ★ Lifecycle management works at bucket level, enables to perform an action to objects based on rules.
- ★ Actions
  - ★ Transition – Objects are transitioned from one storage class to another.
    - ★ STANDARD or REDUCED REDUNDANCY to STANDARD\_IA.
    - ★ STANDARD to GLACIER.
    - ★ Objects must be stored for at least 30 days in the current storage class before transitioning.
  - ★ Expiration – Objects are expired and deleted.



IntelliPaat



# Storage Class Analysis



IntelliPaat

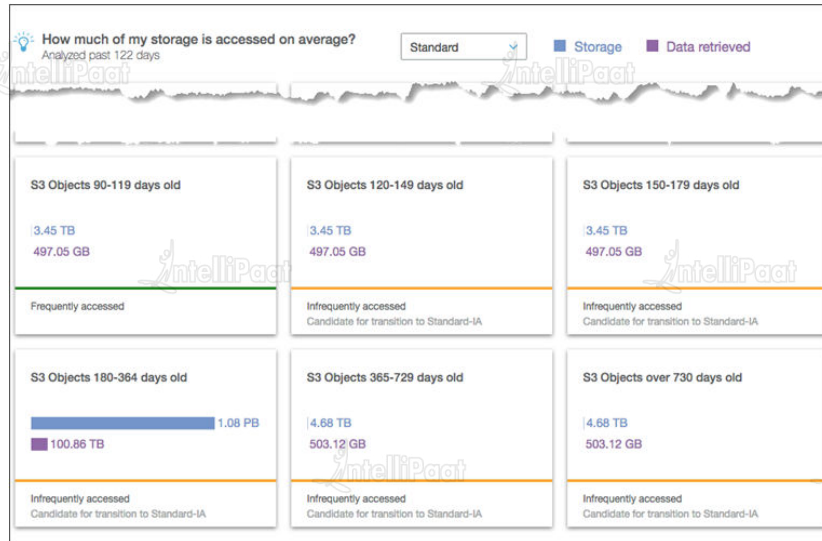
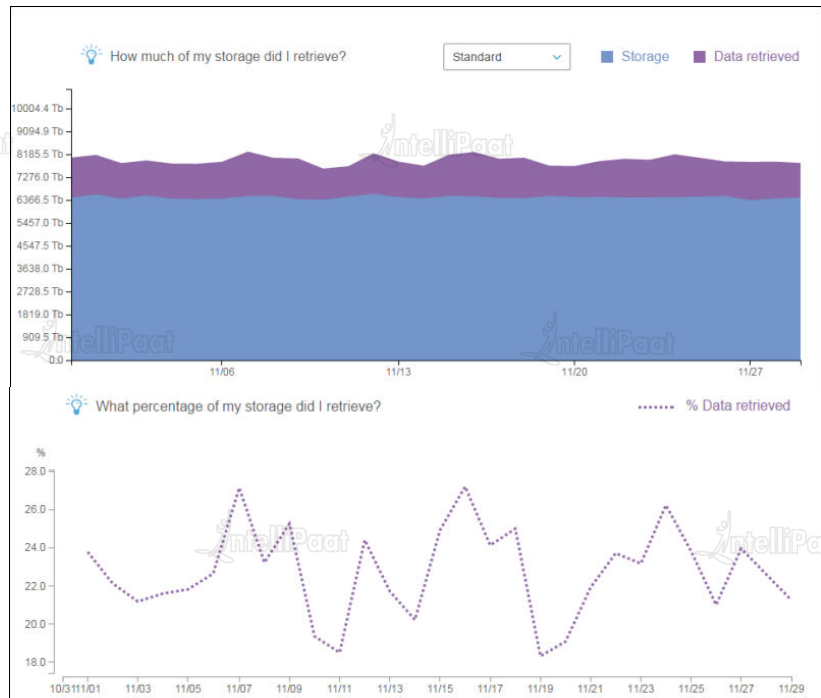


# Storage Class Analysis

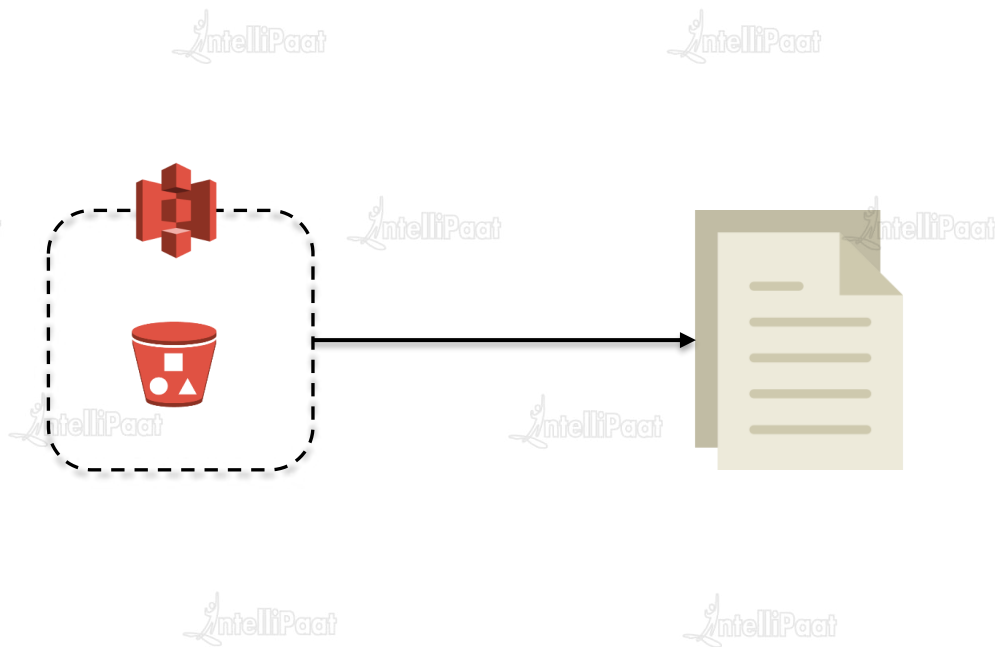


- ★ Provides storage access patterns which can help to decide when data/objects should be transitioned.
- ★ Maximum 1000 storage class filtered analysis per bucket.
- ★ Analysis patterns
  - ★ Analyze entire content of a bucket.
  - ★ Analyze objects grouped by tags or prefixes.
- ★ Storage class analysis observes the access patterns of a filtered object data set for 30 days or longer to gather enough information for the analysis, a message is displayed in the Amazon S3 console.
  - ★ How much of data was retrieved out of total storage.
  - ★ What percentage of storage was retrieved.
  - ★ How much of storage is infrequently accessed.
  - ★ Data can be exported for future analysis.

# Storage Class Analysis



# Inventory



- ★ Inventory provides report for Objects and its metadata on a daily or weekly basis in a comma separated output file.
- ★ Metadata output is configurable.
- ★ Source bucket – Which inventory is created for.
- ★ Destination bucket – Where inventory is stored in.



# Demo 4: Lifecycle Management

---



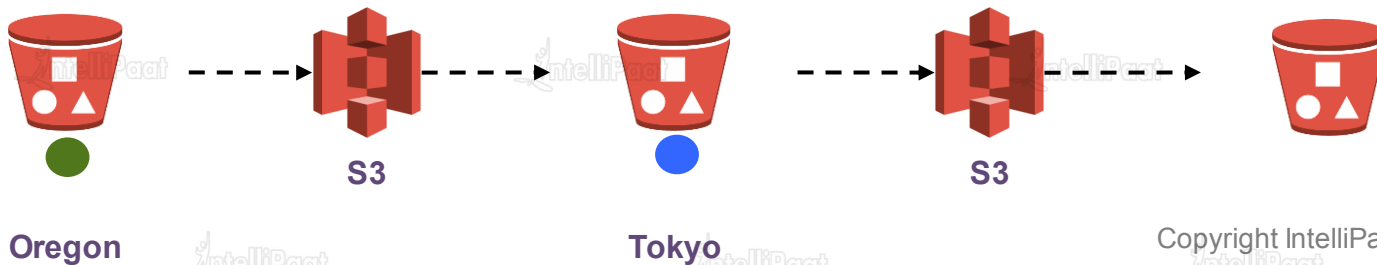
- Setup Lifecycle management for only a few objects in a bucket.
- Setup Lifecycle management for objects with tag “Training”.

# Cross-Region Replication

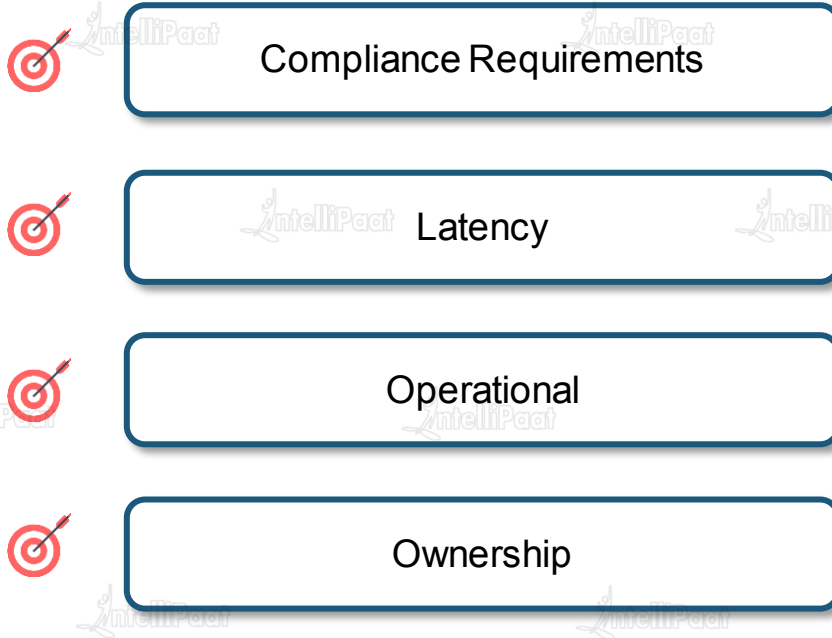
# Cross-Region Replication



- ★ Cross-region replication - Automatic asynchronous replication of objects to a different region. Subset of objects can also be replicated using prefix matches.
- ★ Versioning should be enabled for CRR to work.
- ★ Source bucket or its objects can be replicated to only one target bucket.
- ★ Deletion of specific object version is not replicated over to the other region.
- ★ Existing objects of a bucket are not replicated (if replication is enabled later on).
- ★ Lifecycle management actions are not replicated.
- ★ Replicated objects are not replicated to other region.



# Cross-Region Replication





IntelliPaat  
IntelliPaat



# Data Encryption



IntelliPaat



# Data Encryption

Encryption can be done for data in transit and data at rest when.

## Server side encryption

- ★ S3 encrypts data at the object level as it writes to disks in its data centers and decrypts it when accessed. "x-amz-server-side-encryption"
- ★ SSE-S3 □ x-amz-server-side-encryption: AES-256
- ★ SSE-KMS □ x-amz-server-side-encryption-aws-kms-key-id: <kms\_key\_id>
- ★ SSE-C □ customer algorithm, customer key and customer key MD are passed.

## Client side encryption

- ★ Client-side encryption refers to encrypting data before sending it to Amazon S3. Following two options are available for using data encryption keys
- ★ AWS KMS-managed customer master key
- ★ client-side master key



Client

^#!~ +



AWS Data Center

ABC

'%\$#@

# Server access logging

# Server access logging

---



- Access logging enables to track requests at bucket level. Access logs are stored in separate buckets.
- Access log format
  - Bucket owner – Owner of the source bucket.
  - Bucket name - Name of the bucket that the request was processed against.
  - Time – Time at which request was received.
  - Remote IP – IP address of the requestor.
  - Requester – ID of the requestor.
  - Operation – REST.*http\_method.resource\_type*.
  - Key – Object key in URL.
  - Request-URI - Request-URI part of the HTTP request message.
  - HTTP Status, Error Code & Bytes Sent.
  - Object Size – Total size of the object in bytes.
  - Total Time - Measured in ms from the time request is received to the time that the last byte of the response is sent.
  - Turn-around Time – Number of milliseconds that S3 spent processing the request.
  - Referrer, User-agent, Version ID.



# Demo 6: Website hosting



- Create a bucket - <your-name>.s3.static.wesite.bucket (or whatever is suitable for your need).
- Upload your HTML page as an object.
- Create sub-folder1 in the bucket.
- Upload another HTML document in the sub-folder1. Name of this HTML document should be same as that of the one in the bucket
- Create sub-folder2 inside sub-folder1 and upload an HTML document with the same name as earlier.
- Access the root page and all the subpages using S3 endpoint.
- Create one more bucket. Make this bucket to hold all logs for the bucket created in step 1.
- Configure redirect so that requests to your main page are redirected to /sub-folder1/sub-folder2.
- Integrate S3 website hosting with a custom Domain name using Route53.



IntelliPaat



# Connect using VPC endpoint



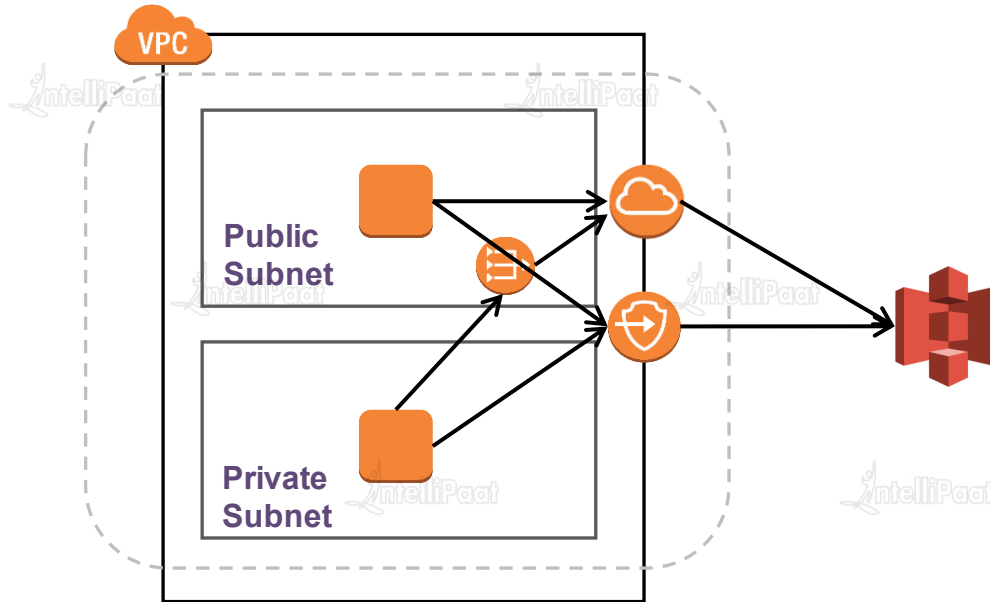
IntelliPaat



# Connect using VPC endpoint



Connect to S3 from EC2 instances in Private Subnets, so that traffic never leaves Amazon's N/W.



# Connect using VPC endpoint



- <https://aws.amazon.com/s3/pricing/>
- Storage
- Standard
  - \$0.023/GB for first 50 TB/month
  - \$0.022/GB for next 450 TB/month
  - \$0.021/GB for next 500 TB/month
- Standard – IA: \$0.0125 per GB
- Glacier - \$0.004 per GB
- Requests
  - PUT, COPY, POST, LIST - \$0.005 per 1000 requests
  - GET and All Other - \$0.0004 per 1000 requests

Total Storage – 750 TB

$$(50 \times 0.023 \times 1000) + (450 \times 0.022 \times 1000) + (250 \times 0.021 \times 1000) = \$16300$$

150 million GET requests =

$$(150,000,000 / 1000) \times \$0.0004 = \$6$$

$$500,000 \text{ PUT requests} = (500000 / 1000) \times \$0.005 = \$2.5$$



IntelliPaat



# S3 Pricing (us-east-1)



IntelliPaat



# S3 Pricing (us-east-1)



## Data Transfer

Data Transfer IN from ANYWHERE is free

Data Transfer OUT to Internet

First 1 GB/month – FREE

Next 10 TB/month - \$0.09 per GB

Next 40 TB/month - \$0.085 per GB

Next 100 TB/month - \$0.07 per GB

More than 150 TB/month - \$0.05 per GB

**Download per month – 80 TB**

$$(10 * 0.09 * 1000) + (40 * 0.085 * 1000) + (29 * 0.07 * 1000) = \$6330$$



IntelliPaat



# Quiz



IntelliPaat



## 1. Default limitation of Bucket in each account?

A. 100

B. 200

C. 1000

D. No Limitation



## 2. Objects in the Bucket can be up to what size?

A. 1 TiB

B. 100 TiB

C. 1 GiB

D. 5 TiB

**3. After enabling Cross Zone replication older data replicated to destination within how much time?**

A. 5 mins

B. Depends on Load

C. Older data never gets replicated

D. 1 hour

## 4. How large can a single archive be?

A. 1 TiB

B. 100 TiB

C. 1 GiB

D. 40 TiB



**India : +91-7847955955**

**US : 1-800-216-8930 (TOLL FREE)**



**[sales@intellipaate.com](mailto:sales@intellipaate.com)**



**24X7 Chat with our Course Advisor**