# **Amazon S3: Custom Cost-Optimized Storage Strategy Implementation**

## Objective

To implement a customized, cost-optimized Amazon S3 storage lifecycle for a dataset using manual tagging, intelligent tiering, lifecycle policies, and cost monitoring mechanisms via the AWS Console.

## 1. S3 Bucket Creation and Object Upload

Bucket Name: rishabh-imdb-reviews

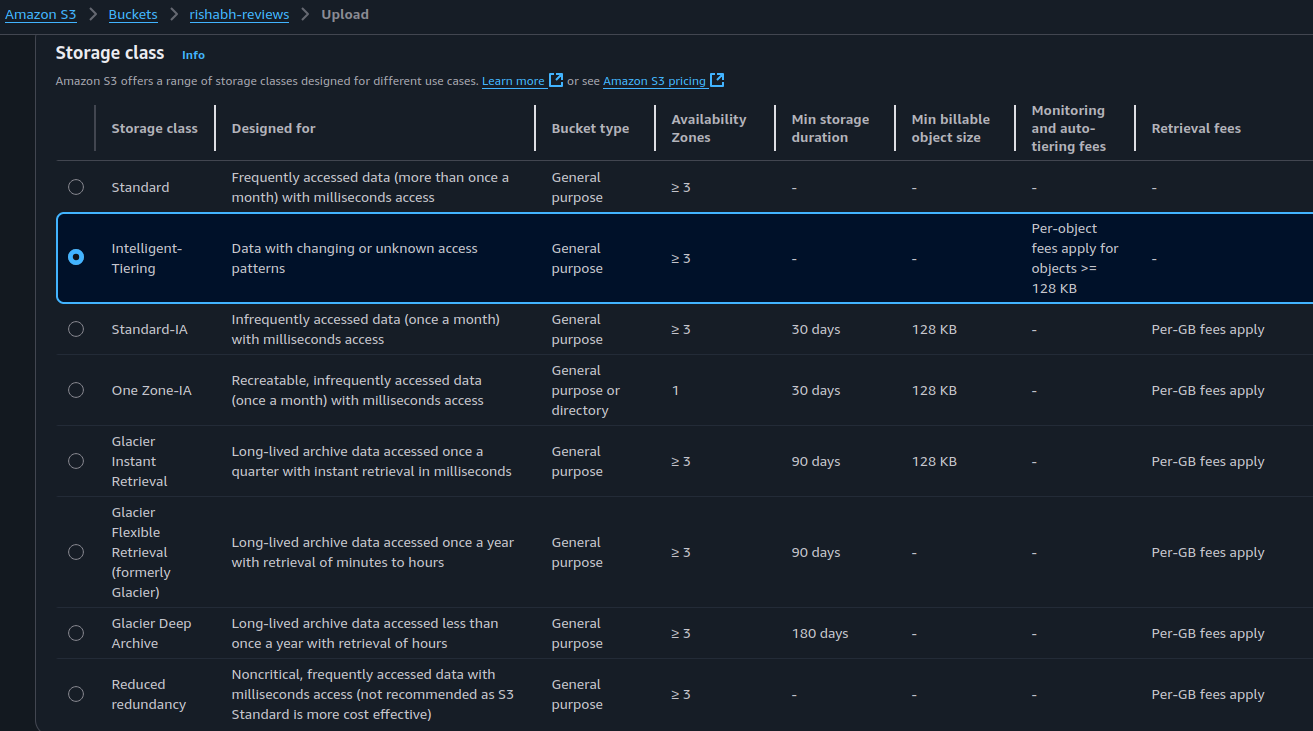
Region: ap-south-1

Versioning: Disabled (can be enabled later for retention)

Encryption: Enabled (SSE-S3)

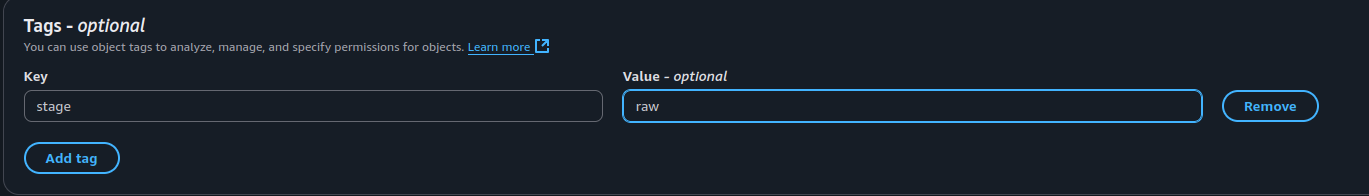
Block Public Access: All enabled (default and recommended)

Storage Class at Upload: Intelligent-Tiering



**Tags Applied on Upload:**

* Key: stage



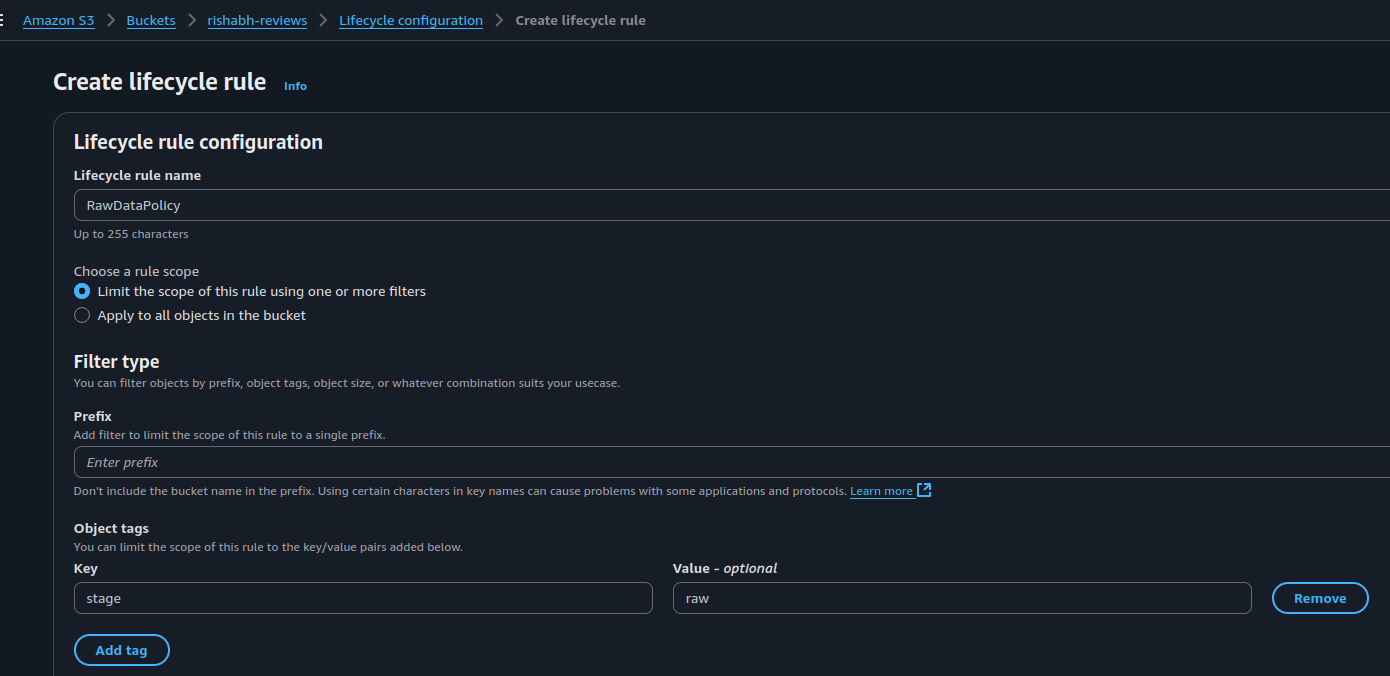
* Value: raw

**Dataset Used:**

* Name: IMDB Dataset of 50K Movie Reviews
* Source: https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
* File: IMDB Dataset.csv

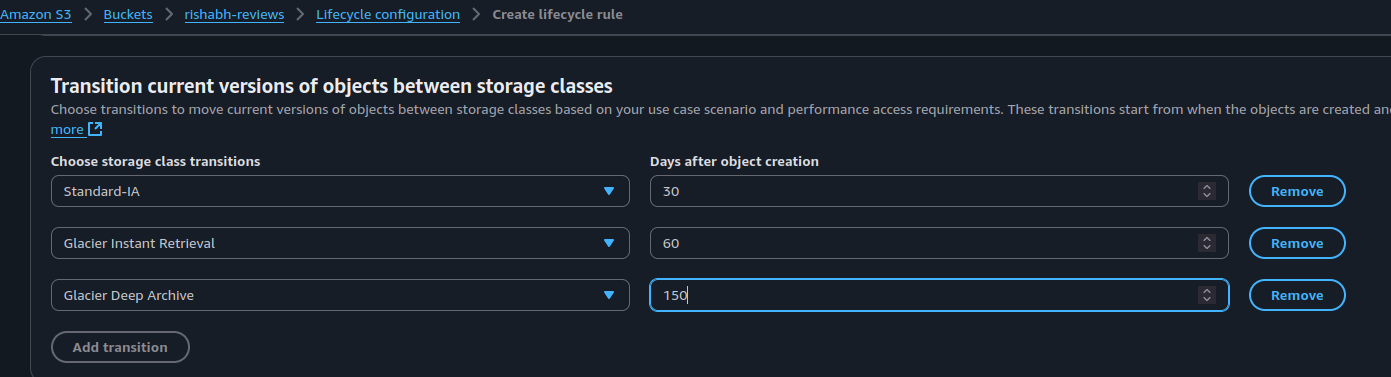
## 2. Lifecycle Rule Configuration

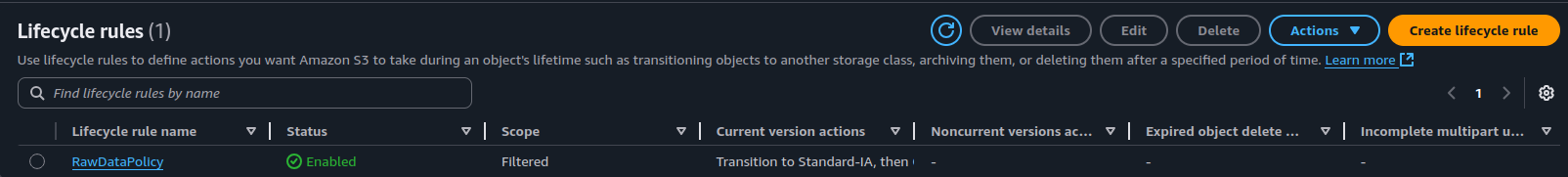
Separate lifecycle policies were created for the stage=raw tag:



**Rule 1: Raw Data (stage=raw)**

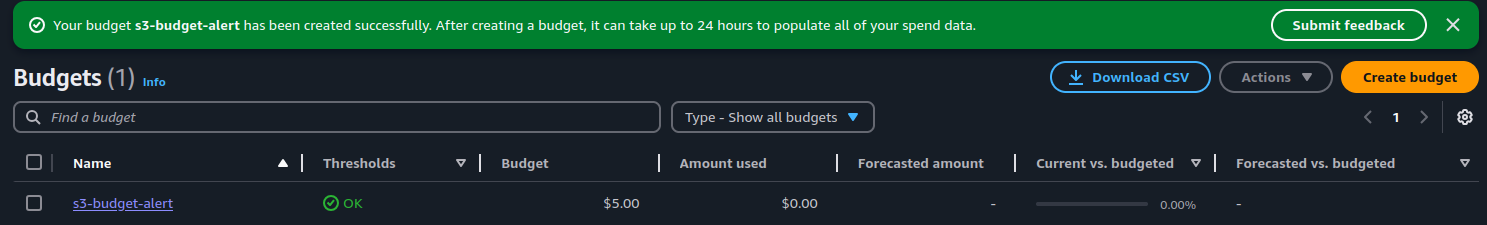
* Transition to **Standard-IA** after 30 days
* Transition to **Glacier Instant Retrieval** after 60 days
* Transition to **Glacier Deep Archive** after 150 days



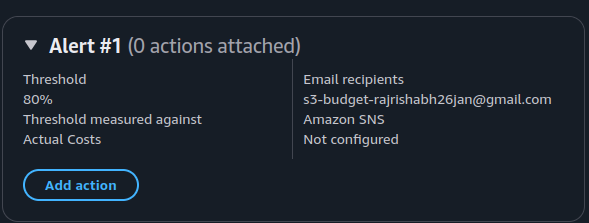


**NOTE:** Rules for stage=processed and stage=archive can be created similarly as data progresses through stages.

## **3. Cost Budget Configuration**

****

* Tool: AWS Budgets (via Billing Console)
* Budget Type: Monthly cost budget
* Service Scope: Amazon S3
* Alert Threshold: 80% of budget limit
* Notifications: Email alerts configured



## **Summary**

**Bucket & Upload:** Created rishabh-imdb-reviews in ap-south-1 and uploaded IMDB dataset object tagged with stage=raw using Intelligent-Tiering.

**Lifecycle Policies:** Defined one tag-based rule to transition data:

raw → Standard-IA (30d) → Glacier Instant Retrieval (60d) → Glacier Deep Archive (150d)

**Cost Monitoring:** Configured a monthly S3 budget with 80% alert threshold via email.

This strategy ensures automated, tag-driven data tiering and cost visibility for long-term S3 storage, while remaining within AWS Free Tier constraints.