WHIZLABS

Home / AWS / Guided Lab / AWS S3 Multipart Upload using AWS CLI

AWS S3 Multipart Upload using AWS CLI

Level: **Advanced**

Amazon EC2 | Amazon S3 | AWS CLI | Amazon Web Services

Required Points	10
Lab Duration	01:00:00
Average Start time	Less than a minute

Start Lab →

Need help?

- How to use Hands on Lab
- Troubleshooting Lab
- FAQs

Submit Feedback | Share

Lab Overview

Cloud Developer, Cloud Administrator

Storage, Developer Tools

Lab Details

1. This Lab walks you through the steps on how to upload a file to an S3 bucket using multipart uploading.

2. Duration: **01:00:00 Hrs**

3. AWS Region: **US East (N. Virginia) us-east-1**

Introduction

What is S3?

- S3 stands for Simple Storage Service.
- It provides object storage through a web service interface.
- Each object is stored as a file with its metadata included and is given an ID number.
- Objects uploaded to S3 are stored in containers called **Buckets**, whose names are **unique** and they organize the Amazon S3 namespace at the highest level.
- These buckets are region specific.
- You can assign permissions to these buckets, in order to provide access or restrict data transaction.
- Applications use this ID number to access an object.
- Developers can access an object via a REST API.
- Supports upload of objects.
- Uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network.
- Designed for storing online backup and archiving of data and applications on AWS.
- It's mainly designed with the minimal features in order to create web-scale computing in an easy way.
- Storage classes provided are:
    - Standard
    - Standard_IA i.e., Standard Infrequent Access
    - Intelligent_Tiering
    - OneZone_IA
    - Glacier
    - Deep_Archive
    - RRS i.e., Reduced Redundancy Storage (Not recommended by AWS)

- Data access is provided through S3 console, which is a simple web-based interface.
- Data stored can be either Public or Private based on user requirement.
- Data stored can be encrypted.
- We can define life-cycle policies which can help in automation of data transfer, retention and deletion.
- Amazon Athena can be used to **query** S3 data as per demand.

Uploading and copying objects using multipart upload

- Multipart upload allows you to upload a single object as a set of parts.
- Each part is a contiguous portion of the object's data.
- You can upload these object parts independently and in any order.
- If transmission of any part fails, you can retransmit that part without affecting other parts.
- After all parts of your object are uploaded, Amazon S3 assembles these parts and creates the object.
- When your object size reaches 100 MB, you should consider using multipart uploads instead of uploading the object in a single operation.

When to use multipart upload

- If you're uploading large objects over a stable high-bandwidth network, use multipart upload to maximize the use of your available bandwidth by uploading object parts in parallel for multi-threaded performance.
- If you're uploading over a spotty network, use multipart upload to increase resiliency to network errors by avoiding upload restarts.
- When using multipart upload, you need to retry uploading only parts that are interrupted during the upload. You don't need to restart uploading your object from the beginning.

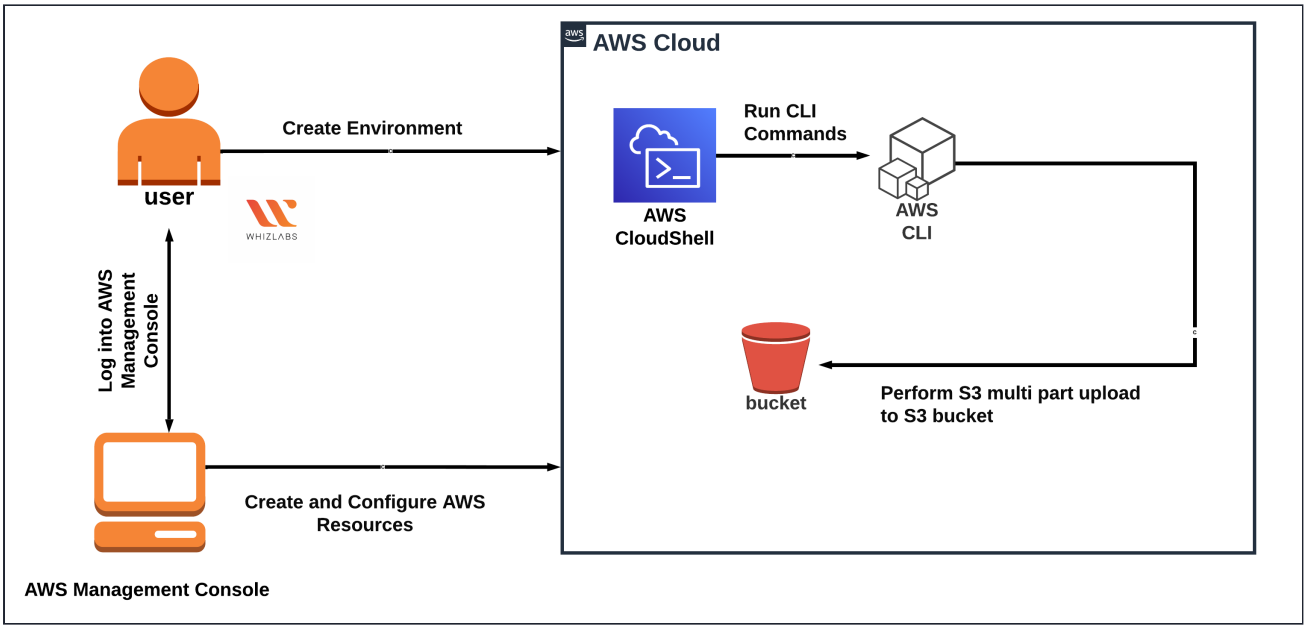
Multipart upload process

- Multipart upload is a three-step process:
    - **Multipart upload initiation :** When you send a request to initiate a multipart upload, Amazon S3 returns a response with an upload ID, which is a unique identifier for your multipart upload. You must include this upload ID

whenever you upload parts, list the parts, complete an upload, or stop an upload. If you want to provide any metadata describing the object being uploaded, you must provide it in the request to initiate multipart upload.

- **Parts upload :** When uploading a part, in addition to the upload ID, you must specify a part number. You can choose any part number between 1 and 10,000. A part number uniquely identifies a part and its position in the object you are uploading.
- **Multipart upload completion :** When you complete a multipart upload, Amazon S3 creates an object by concatenating the parts in ascending order based on the part number. If any object metadata was provided in the initiate multipart upload request, Amazon S3 associates that metadata with the object. After a successful complete request, the parts no longer exist.

Architecture Diagram



Task Details

1. Sign in to AWS Management Console
2. Create an S3 bucket
3. Create an Environment in CloudShell.
4. Copy a video file from s3 bucket using CLI

5. View the original file
6. Split the original file
7. Create a Multipart upload
8. Uploading the file chunks
9. Create a Multipart JSON file
10. Complete the Multipart Upload
11. View the file in the S3 Bucket
12. Validation of the lab

Launching Lab Environment

1. To launch the lab environment, Click on the **Start Lab** button.
2. Please wait until the cloud environment is provisioned. It will take less than a minute to provision.
3. Once the Lab is started, you will be provided with **IAM user name**, **Password**, **Access Key**, and **Secret Access Key**.

Note : You can only start one lab at any given time