AWS Storage

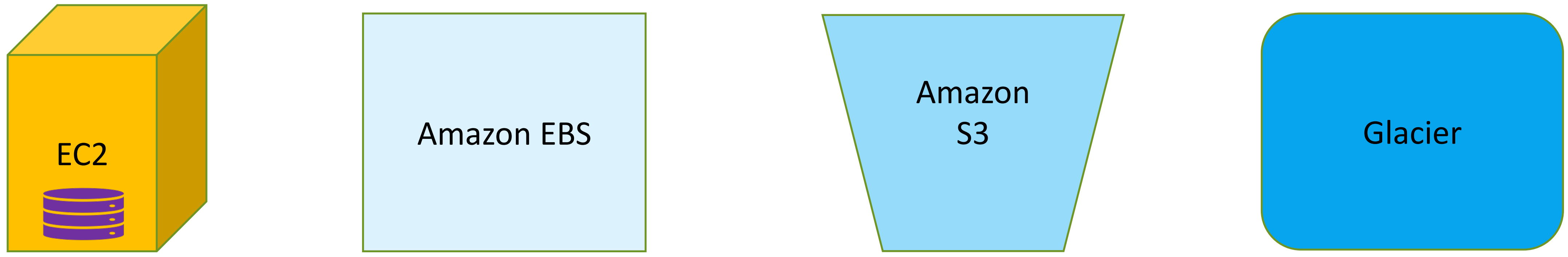
Analytics Tensor

Mahesh KC

mahesh.kc@analyticstensor.com

https://analyticstensor.com

Storage in AWS



EC2

Amazon EBS

Amazon S3

Glacier

Instance Store

- Built-in to EC2
- Cost included in EC2
- High IOPs
- Ephemeral
- No Snapshots

Block Storage

- Independent of machine
- $/GB/Month
- Up to 16 TB
- Durable (replicates data in AZ's)
- Snapshots

Object Storage

- Object-Storage
- $0.023/GB/Month
- Write once read many
- Storage for the internet
- All transfers via HTTPS

Cold Storage

- Archival Storage
- $0.004/GB/Month
- Transition from S3
- Slow retrieval

Ephemeral stores are deleted when instance is stopped or terminated.

AWS Storage Gateway

The AWS Storage Gateway is a service connecting an on-premises software appliance with cloud-based storage to provide seamless and secure integration between an organization's on-premises IT environment and AWS's storage infrastructure. It is installed as VM on VMware ESXI or Microsoft Hyper-V. Stores all data encrypted in Amazon S3 or Amazon Glacier. It has three types of configuration:

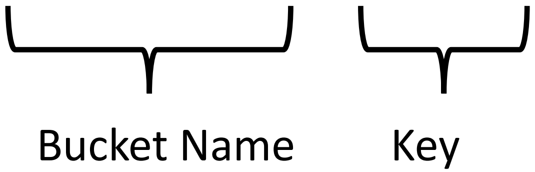
- Gateway-Cached Volumes: Stores majority of data in S3 and stores frequently-accessed data locally.
- Gateway-Stored Volumes: Stores all the data locally. It takes point-in-time snapshots to S3.
- Gateway-Virtual Tape Library: Exposes iSCSI interface. It looks like tape and each tape is stored in Amazon S3 or Amazon Glacier.

Simple Storage Service (S3)

It is an object storage. It provides fault tolerance and is highly available since cluster spans over entire region. It has no filesystem. It has flat hierarchy. i.e. Bucket and Object. For example.

https://s3-us-east-1.amazonaws.com/analyticstensor/files/1.txt

Bucket Name: analyticstensor
Key: files/1.txt



The bucket name need to be globally unique among all of the S3. No one on the world across every region would be able to use that bucket name so it must be globally unique. There is not bucket size and number of objects limit inside the bucket. There is limit of size in individual objects i.e. 5TB. The upload limit is 5GB but multipart upload can be used to upload larger objects. We can also specify server side encryption (SSE) by specifying flag when uploading the file. It use AES 256 encryption.

What is difference between Object Storage and Block Storage? (Self Assignment)

Demo: Create Bucket and Objects. Give access to object for world.

Bucket Security with Resource Policies

It is similar as IAM policies but it is applied at resource level. It specifies a principal. i.e Who is allowed to perform particular actions or who is denied for action. We can specify permission to:

- Another account
- IAM user, group, role
- AWS Service
- Anonymous (to anyone in the world)

Example of Bucket Policy

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": "*",
      "Action": [
        "s3:GetObject"
      "Resource": [
        "arn:aws:s3:::analyticstensor/*"
      ]
    }
  ]
}
```

Public read policy to s3 bucket

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
         "AWS ": "arn:aws:iam::512345678:root "
     },
      "Action": [
         "s3:GetObject”,
        "s3:PutObject”,
        "s3:ListObject"
      ],
      "Resource": [
           "arn:aws:s3:::analyticstensor",
           "arn:aws:s3:::analyticstensor/*"
      ]
    }
  ]
}
```

← another account

Cross-Account policy to s3 bucket

Demo: Apply Bucket Policy on Object

Amazon Glacier

- Used for Cold Storage.
- Pricing:
- Used for Archival storage.
- Write once ready rarely.
- Used to store logs or other files for future retrieval or data retention policy.
- In S3, we can download objects at anytime but in Glacier we need to do retrieval request which take around 3-5 hour wait time for file to be available. Afterwards the file/object is send to S3, provided with link for download.
- For write archives, we can transition from S3 or do direct upload.

Lifecyle Rules: It is a way to manage storage class. We can manage in:

- Standard
- Standard_(IA (Infrequent Access)
- Glacier.

We can archive files to Glacier Storage Class directly from S3. For example, if we want to move all object from S3 with objects prefix logs/ then we can specify a lifecyle rule to move into Glacier. We can specify N days to move into Glacier and N days to permanently delete. This is an automated way to store files in S3 then move to Glacier and delete afterward.

Demo: Adding Lifecyle Rules in S3 buckets.

Instance Store Volumes

Instance Store Volumes are volumes that comes with EC2. It is build-in to EC2. We are not paying any extra since it is built-in with E2 and price will be based on runtime of EC2 instances. It has very high IOPs. But there are some trade off:

- It is Ephemeral. i.e. Stores are deleted when EC2 instance is terminated.
- No Snapshots.

For example,

m3.medium: SSD 1 x 4 GB

m3.large: SSD 1 x 32 GB

m3.xlarge: SSD 2 x 40 GB

d2.8xlarge: HDD 24 x 2TB

hi1.4xlarge: SSD 2 x 1TB

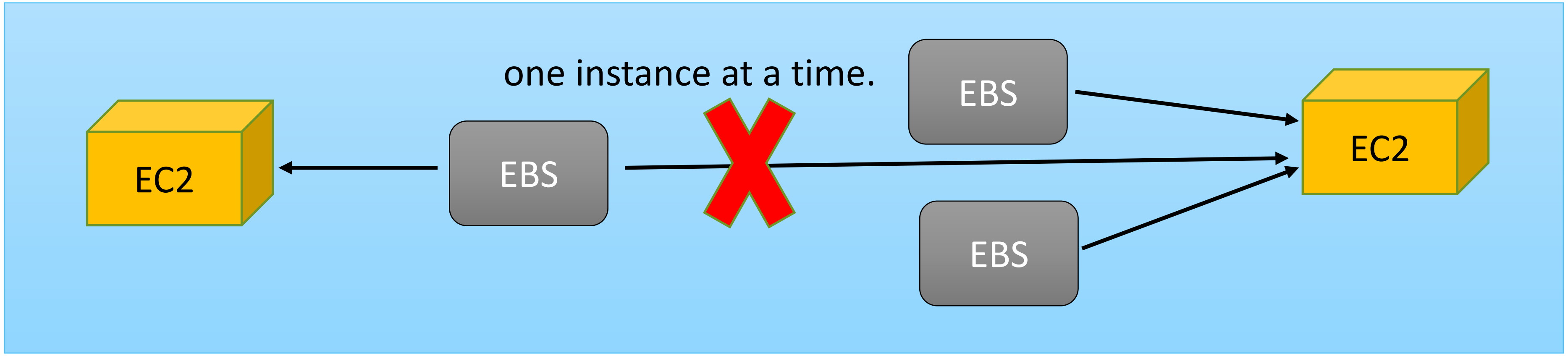
Note: t1, t2, m4 and c4 series have no instance store volumes.

Amazon Elastic Block Store (EBS)

- Data is independent from instance.
- Pay for whatever we provisioned storage. For e.g. if we specify 1 TB, then we need to pay for 1 TB.
- It exists in single AZ.
- It can be encrypted either in OS level encryption or EBS built-in encryption mechanism.

We can think EBS as external hard drive on USB drive. We can attach and detach the EBS. But we can only have one EBS volume attached at one time. We can have multiple EBS attached to one EC2.

Self Reading: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html



one instance at a time.

Amazon Elastic Block Store (EBS) (cont.)

Demo: Creating an EBS Volume

Click on EC2 on services

Under Elastic Block Store, choose Volumes (left side of screen)

Click Create Volume

Choose Volume Type, Size, IOPS, Availability Zone, Encryption(optional)

Click Create

Rename the volume. For e.g. Data Volume.

Mounting EBS Volume on Linux Machine

Choose the EBS volume.

Click Actions, Choose Attach Volume, Choose linux server.

Click Attach. Make sure the volume is mounted.

Login to VM using ssh.

Type: lsblk (it shows the volume i..e xvdf)

Format disk to ext4, Type: sudo mkfs –t ext4 /dev/xvdf

Create a mount point: first create mount point: Type: sudo mkdir –p /mnt/data

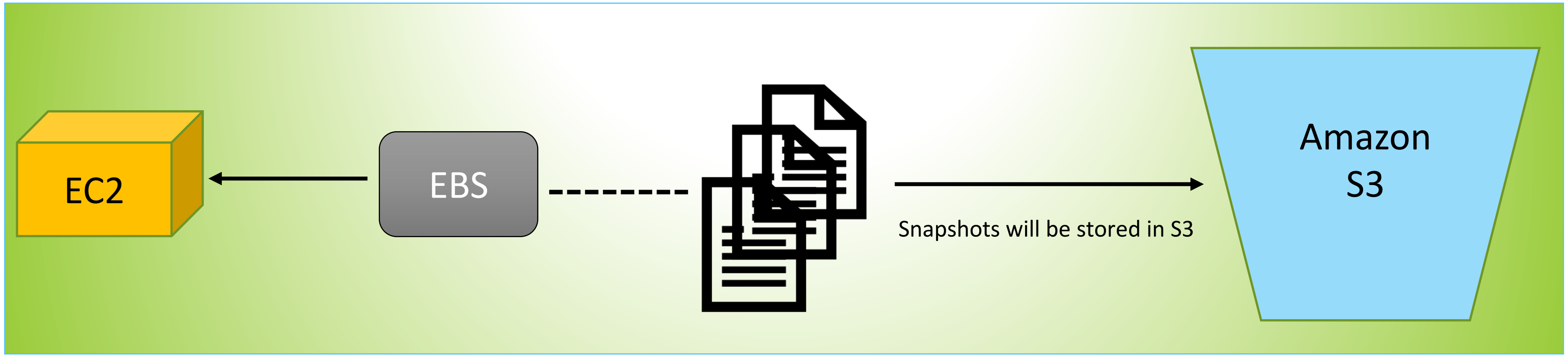
Mount the drive: sudo mount /dev/xvdf /mnt/data

Check: df

Note: To attach to EC2, both must be in same Availability Zone.

Amazon Elastic Block Store Snapshots

- EBS volumes exits in single AZ.
- It is durable to loss of device.
- It is not durable to loss of AZ.
- When we lost the AZ then it will be lost. We can create snapshot which will be stored in S3.
- We can easily copy the snapshot to other region for disaster recovery. For e.g. we can copy from us-east-1 to us-west-1.
- We can also share the snapshot between multiple account. i.e. copy the data from production account to other QA or DEV account. EBS snapshots increase the durability of the data for geographical diversity and helps to create volume if they are lost.


EC2
EBS
Snapshots will be stored in S3
Amazon
S3