Storage

24 May

Storage :

Storage Needs for enterprice apps

- WebApps
  - To store media(images,docs,videos)
- Backups
  - To store backups
- Applications
  - Disks
- IT organisations
  - shared disk

Cloud usage

- think cloud as infra as a software (someone else's hardware and we pay accoding to usage)
- earlier all the data from the computers will be stored in one common server and at the weekend the server backup will be copied to a tape and stored them in Banks as Archives.
- Backup:
  - used for quickly recovering form disasters
  - it should recover quick
  - prices should be moderate
- Archive:
  - used for recovering from disasters
  - it might takes hours or days of time to recover
  - storge costs should be cheaper

Media storage

- media used by websites:(Flipkart website images)
  - these are accessed from web
  - This is evergrowing

Scenario

- whenever you try to access a fliplart or amazon website the images present in the site will loads very fastly this is beacuse of the edge locatins.

- These edge locations are present near to your location through out the world this reduces the latency and load the media quicker.

- This model is called as **Content delivery network**

- where as when you try to access a server in hyd from delhi it will take a while to load the site fully this because of high latency. like QT website

- Gnerally companies use content delivery networks.

- like netflix whenever you try to access the vedios, those will be loaded very quickly because. they are present in nearby edge locations.

Blob Storage

- A storage of any file type and is accessable over https(URL).
- if you are using a windows/linux machine definetly there wil be a harddisk, this OS will expect a file system in harddisk to store data for windows NTFS ...,and the disk which is used for windows can't be used for linux.
- earlier Dropbox came up using the same idea, dropbox uses AWS storage in backend

25 May

Terms and Concepts

- 1 kibi byte = 1024 bytes(1 kib, 1GiB, 1TiB)
- 1 kilo byte = 1000 bytes (1kb, 1GB,1TB) ---> new Metrics
- 1 kilo gram = 1024 grams
- storage industry adopted new metrics but OS dindn't
- 500 GB= 500 * 1024 (incorrect)
- 500GB = 500 * 1000 (correct)
- but our OS still thinks as for 500GB = 500 * 1024, [1000/1024]* 500 = 488.23gb
- this means the the 500gb will be viewed in old OS metrics.
- how would you define the speed of your computer ?
  - speed for your pc will be specified by clock speed
  - how fast your RAM will be known by size of your RAM
- Hard disk types:
  - Magnatic
  - HDD
  - SSD
- how about Disk Speed:
  - Throughputs --> generelly used for HDD
  - IOPS(I/O per second) --> used for SSD's

```
* Throughput: The rate at which data is transferred over a network or processed by
a system, typically measured in megabytes per second (MB/s) or gigabytes per
second (GB/s).
* IOPS (Input/Output Operations Per Second): A performance measurement that
indicates the number of read and write operations a storage device can perform in
one second.
```

- IOPS:
  - definies as number of times you can do read and write operations.

  - the amount of data which is carried out within a second defines the IOPS
  - Note: the amount of data is not always same it might differ EX : with a 100ml jug
- Throughputs:
  - Througput speaks about how much of data you have been trasferred within a second.
  - EX: using 100ml jar you took water out from a bucket 10 times (100*10=1000ml) which speaks about 1000ml in one second
- File System:
  - Every disk needs some filesystem (which organizes the whole storage to be used by operating systems)
  - windows file systems: NTFS
  - Linux File System: ext4.xfs,btrfs,zfs

```
    In windows you can do partitions of disks the devided disk will become as
logic disks you can start using them.

    In Linux it won't work like that while doing partion you need to mention that
to which folder you are going to mount it, from then the data which is there in
that particular folder will be stored in seperate disk.
```

- Network Storage:
  - In some cases we need the disks to be shared across multiple devices, Generally for this NAS, SAN, NFS[network file share]

```
Question : in a company there are some 6 employees to their systems we need to
send a 100gb data?
How companies solve this :
    Companies will have [Network Attached Storage] to the employee system(a
network storage drive is mapped in all the employees system)
```

```
Question : when you shutdown your system do you think that the disk also will
shutdown ?
A. NO , if that is the case your operating system needs start again
    so the disks doesn't have state of shutdown, since the cloud providers will
charge for disks even when the vm is shutdown.
```

Storage on Cloud

- All storage devices on cloud are virtual in nature.
- AWS:
  - Elastic block storage: This is equivalent to physical disk on server and disk is called as EBS volume.
  - For the same purpose they also have instance storage
  - Network Disks:

        - EFS(Elastic File System)
        - FSx:
            - NetAPP

- Azure:

    - Managed Disk: This is equivalent to physical disk on server
    - Unmanaged Disk: This is equivalent to physical disk on server
    - Temp Disk(Local Storage): This is also a disk
    - Network Disks:
        - Azure File Shares
        - Azure NetApp

- Backup of Disks:

    - Types:
        - Full
        - Incremental backups(possible when you have previous day backup)
    - Recovering from Backups is also called as **Restore**
    - In most of cloud, Backup of a disk is reffered as **Snapshot**

26 May

- Disk storage in AWS:
- we know that in AWS wehave regions
    - Regions-> Availability zones -> Data centers -> Racks -> Blade serves.

Question : Q. when we create a server the storage comes along with the same vm or different vm ?

A. No, the disk will come from a different server same Availability zone.

- The disk from a different server called as **Elastic Block Storage** advantage is that even you delete the server the disk will be avilable for you.
- If the disk is coming from the same physical machine called as **Instance storage** when you shutdown the machine all the data will be lost.
- Temporary in cloud says as **ephemeral**
- when you want to be data temporary you can go with Instance storage if you want persistant data go with Elastic block Storage.

EC2 creation (w.r.t disk storage)

- Ec2 instance when create will need disk storage.
- The disk or volume with os has to be an EBS volume
- EBS volume is created on different physical server in the same AZ as EC2.
- If the disk is created from same physical server as EC2 we refer it as instance storage.
- In AWS all ec2 instance will have a disk with os as EBS Volumes. Instance storage is supported on fewec2 instance types.
- INstance store is temporary or ephemeral in nature i.e. when we stop the ec2 instance , data in instancestore is lost.In AWS, we can add additional ebs volumes.
- Instance storage option is not avilable for all the storage types.

VM in Azure

- In Azure it is same as Aws but names will differ, Os disk is called as Managed Disk, and instance storage called as Local storage temp disk
- procedure is same when you create a VM the storage disk will be from other rack/server which is known as Managed disk and the **instance storage also comes with most of the storage types in Azure** instance storage in Azure we call this as Local storage disk(temporary disk)

Do This

```
* Create a vm by selecting a storage type which comes with local storage
* go to resource group where the vm is created and select the storage there you
will find storage as 30GB
* this disk will be there even after you delete the vm
* now connect to this vm
* type `sudo lsblk`(list all the block devices)
* find sda having 30gb and find sdb as 4Gb
* temp disk is mounted to `/mnt`
* do `ls /mnt`
* do ` touch {1..100}.txt` this will create 100 files in that folder
* do the same `cd ~`
* do `mkdir myfile`
* do `cd myfile/`
* do `touch {1..100}.txt`
* now stop the vm and restart it again and check the files you created now you
will find only in `sda` in other folder `/mnt`you can't find them because it is a
temp disk.
```

- Azure Disk Terms
  - OS Disk: disk with os
  - data disk : additonal disks
  - temp disk : this represents ephemeral storage
- In Azure number of data disks is dependent on the size of vm.
- when you are selecting a type of disk size check the data disks count, which tells about how many data disks you can attach to the VM.
- the data in these disks are not encrypted but there is an option we can encrypt them
- the applications which allows to do payments the data need to be encrypted ,the payroll data in the storages will be encrypted using `PCI STANDARDS`
- there is newly luanched thing that we can encrypt our physical disks too, which prevents from data stealing.

28 May

Windows Server

- Create a windows server in aws and login into it and navigate to server manager and check the disks from `files and storage option`
- Now we will try to add one more ebs volume i.e a disk to Ec2 machine

- Now create a volume of 1GB and attach to Ec2 machine

```
 * go to Volumes `create volume`
 * check the location and size and create it, and select the created    volume and click on actions select attach volume and attach to the created VM.
 * Now go to VM and refresh the server manager page you will find a new disk attached to it.
 * Amazon job is ended till here, to use it we needs to below steps
 * right click on the disk and select `Bring Online`
 * now you have two optoions here
    * Right click you will find `create a new volume`
    * select the other disk right click on it you find `extend volume`
* For as of now, we will create it as a new volume, go to disk and right click on it create a new volume and fill the procedure you will find a new volume created in `This PC`.
```

- Now lets increase the size of created ebs volume to 2GB.

```
    Note: Earlier in Amazon whenever you increase the disk size the only way to resize it is to stop the machine and start again, now it is changed exactly 4 years back you can directly use the disk by incresing the disk size
    * Decreasing the EBS volume sizes is not allowed.

    how to increase the disk size:
    * go to volumes, select the vloume created and click on Actions and select `Modify volume` and change the size to 2GB and select modify.
    * done , wait for the staus
    * go to sever manager and refresh it, slect the disk from below options and right click on it and extend volume enter the maximun disk size shows there.
    * now go to `This PC` and verify you will see a increased disk to 2GB

    Note : whenever you delete the vm you need to delete the 2GB volume manually. newly attached disks won't get deleted along with vm
```

Linux Server

- Create a Linux VM and while creating, go to storage section and add one more disk of size 1GB.
- select yes for delete on termination.

```
* connect to VM from powershell
* do `sudo lsblk` to list out the list of disks
* if you observe XVDA is of 30 GB and XVDB is of 1GB
* Now we will add one more disk of 1 GB to the same VM, go to volume as we did earlier attach it to linux VM. selcct (/sdd) while selecting type
* execute the same command agian to check the changes
* extend the volume for one of the disk and try to observe the changes
* now the newly created disks are not attached to any folders, either we can
```

```
partion the disk or we can mount the disk to some folder
* now we will try to mount the disk to some folder
* create two folders `sudo mkdir /tools` , `sudo mkdir /projects`
* In linux whenever you want to format the disk, that means creating the
filesystem, you can any filesystem.
* For tools folder:
    * do `sudo lsblk`
    * do `sudo mkfs -t ext4 /dev/xvdb`
    * do `sudo mount /dev/xvdb /tools
    * do ` df -h`
* For projects folder:
    * do ` sudo mkfs -t xfs /dev/xvdd`
    * do `sudo mount /dev/xvdd /projects`
* problem with the linux is if you restart you system it will forget about the
mounts
* to make it remember we need to below using `fstab`
    * `sudo blkid` and take a copy of the id's for the new disks.
    *  now open the fstab ` sudo nano /etc/fstab` nano is like your notepad
    * you need to write there in below format
        * ` UUID= _id_ /projects xfs defaults 00`
    * go to the fstab and paste them in fstab and give ctrl+x and save.
    * by doing this it will remember the mounts.
```

About Linux

```
* Linux is having something called as `LINUX KERNAL` it is like a core OS.
* this Kernal is used by all the linux based mahines by adding some additional
features and distributed with different names like
Ubuntu,Kali,Debian,redhat,fedora..etc
```

29 May

Azure VM Creation(w.r.t disks)

- follow for pictorial documentation (quickkstart with keyword ex:azure windows server quick start)
-

Windows server

- Create a windows server in azure with two disks

```
    shortcut to connect with server
    give win + R and enter `mstsc -v _ipaddress_`
```

- bring the additonal disk online from server manager

- go back to server in azure ,go to disks from the options and from there select again `size + operations`select the newly addded disk and extend the storage to 8GB.
- now go to server manager and select the disk you will get an option displayed as Volumes, there select the disk and extend volume give maximun it is showing there.
- **reducing the size of disk is not possible**

Linux Server

- Create a linux VM add additional disk to it(use azure quickstart docs for reference)
- connect to the created server through poweshell and do `lsblk` and verify the three disks are present.
- we will mount the additonal disk to some folder
- `sudo mkfs -t ext4 /dev/sdc`
- `sudo mkdir /projects`
- `sudo mount /dev/sdc /projects/`
- `df -h`
- `susdo blkid` copy block id's `UUID`
- `sudo vi /etc/fstab`
- paste the id in thid command `UUID=_id_ /projects ext4 defaults 0 0`
- if you didn't do fstab whenver the system restarts it will forget about the mounts.

30 May

Disk Types and Pricing

AWS EBS volumes

- In AWS if you want to create a disk below given are the types we have
- EBS Volume Types:
  - gp2
    - SSD
    - Used for general purpose
      - dev *test envi
    - Older generation
    - performance
      - MIN :300 IOPS
      - Max : 3000 IOPS
      - on avg you get 3 IOPS/GB.
  - gp3
    - SSD
    - new generation of gp2
      - cheaper in pricing
      - Better Speed
    - Performance
      - Min : 3000 IOPS

        - Max : 16000 IOPS

  - Provisioned IOPS

    - SSD
    - Used for performance intensive scenarios
    - Performance
      - MIn: 100 IOPS
      - Max: 100000 IOPS (used for disk intensive applications)

  - Throughput Optimized HDD

    - HDD - older generation
    - Min: 125 GiB
    - Max: 16384 GiB

  - Cold HDD -> HDD

  - Magnetic - cheaper and extreamly slow in nature

- **Open AWS pricing calculator**

- estimate storages prices usig pricing calculator

- taking Mumbai as a region

  - Storage size :10 TB
    - gp2:
      - cost : 1167.36 $
    - gp3:
      - cost : 933 $
    - provisioned(io2)
      - cost : 1545 $
      - Iops : 3000
    - cold hdd:
      - cost : 178.18 $
    - Throughputs:
      - cost : 522 $
  - Storage size =10 TB (max speed)
    - gp2: no change (can't set speeds)
    - cold hdd: no change
    - Throughput optimized HDD: no change
    - gp3:
      - cost: 1007 $
      - iops: 1600
    - io2:
      - cost: 11438 $
      - iops: 256000

- HDD's can't be used(waste of memory) for OS i.e root volumes(where the OS exists)

Azure Disk Types

- SSD:
  - standard SSD
    - IOPS :Range(500-6000)
    - size(1-32767)GB
  - premium SSD
    - IOPS: Range(120-20000)
    - size(1-32767)GB
  - premium SSD v2
    - IOPS: Range(80,000)
    - size(1-65536)GB
  - Ultra SSD
    - size: 4-65536 *IOPS: 1024-307200
- HDD:
  - Standard HDD
    - IOPS: Range(500-2000)
    - size(1-32767)GB

Open Azure pricing calculator

- estimate storages prices usig pricing calculator
- search for `Managed disks` and take `East US` as a region
- Storage size : 8 TB
  - Standard HDD:
    - cost: *Standard SSD:
  - Premium SSD:
  - Premium SSDv2:
    - IOPS: 3000
      - cost: 657$
    - IOPS:20000
      - cost: 744$
    - IOPS: 80000
      - cost: 1051 $
  - Ultra SSD:
    - IOPS: 8192
      - cost:1431$
    - IOPS: 400000
      - cost: 21382 $
- In Azure ultra disk can't be used as OS disks.
- It is better to with
  - if Azure: premium v2
  - if AWS: gp3

Backups of EBS volume is called as Snapshot

- Backup of EBS Volume is called as snapshot

- Snapshots can be taken
  - Manually
  - automated
- Using snapshots of root volumes we can create Amazon machine images and for all the other volumes types can be used to create volumes.
- In AWS **EBS** belogs to an availability zone and snapshot belongs to a **Region**. if you need a disk to be created in other region, copy snapshot to other region
- Snapshot can be shared with other aws accounts as well.

Experiment with manual snapshots

- create a ubuntu server with an additional disk of 2gb
- connect to server using powershell
- mount the 2gb volume to some folder
- `sudo mkfs -t ext4 /dev/xvdb`
- `sudo mkdir /app`
- `sudo mount /dev/xvdb /app`
- download `nopcommerce` to `/app` folder
- `cd /app`
- `sudo -i`
- `cd /app`
- `wget _link_` and do `ls` and do `exit`
- In the volume add some data post formatting and mounting.

```
    * Lets create a manual snapshot of this volume.

    Go to AWS -> EC2 instance select the instance and form below blade select
storage and find 2 disks one with 8 and other with 2gb.

    * Taking backup of 2gb disk:
    --------------------------------
        * click on the `volume id`, it will take you to that volume there check
the box for the volume and click on `actions` in actions select `create snapshot`.
        * provide the `description` like `app snapshot`
        * check the status of your snapshot
        * you can find an option of snapshot inside of elastic block store. you
can find all snapshots there.
        * Now go to `VM` and `terminate` it, wait till the vm is terminated
        * now go to the `volumes` and delete the 2gb volume
        * Now create a EC2 machine without any additional disks
        * try creating a volume from snapshot in the same availability zone where
EC2 is created.
        * go to snapshots and select it, click on `actions` and select create
volume from snapshot.
        * select type`gp3` and region `same where your ec2 present`
        * now go to volume and find the volume is created and status as `avilable`
        * now attach volume to the newly created ec2 instance, and connect to the
machine from powershell.
        * SSH into Ec2 and list the block devices `sudo lsblk`
        * now don't try to format the disk by using `mkfs` command because it will
erase everything from the disk.
        * `sudo blkid` and find the file system created already.
        * `sudo mkdir /app`
        * `sudo mount /dev/xvdd /app`
        * `sudo ls /app` here you can find the files.
        * go to `snapshots` and select the snapshot and click on actions and
provide the region `where you want to move the snapshot`. through this way you can
move the snapshots to other regions.
```

June 2

File shares

- Network storage :

```

    You might observed an extra disk with a green pipe symbol in a company
provided laptop which is a Network Storage Drive.
*   This Drives will be connected to the same network where your servers
    are connected.
    Companies like : Netapp, EMC2, PureStorage.

```

File Storage on AWS

- AWS provides two solutions for this
  - EFS(Elastic file system)
    - works in AWS Networks (vpc) only on linux systems.
  - FSx:
    - AWS has launched this service for third party storage solutions
    - FSX for Windows share
    - FSX for NetApp
    - ...
- EFS Modes
  - Elastic(go fot this case :you might not using the data every second)
  - Provisioned throughput(through out the month you will get the same speed)
-

Create an EFS and mount it on two linux instances

- EFS is storage in aws network (vpc)
- EFS requires a security group (open to every one in same network)

```

* Practical:
-------------

* Go to Aws and open VPC, don't create any network there will be a network already
```

```
created in every region `Default`
* take the IP address of networks , open security groups by scrolling down a bit,
create a new security group.
* give name as `EFS` and for below field also EFS.
*  create an inbound rule of type `All Traffic` and source select custom and
provide the `VPC ip address`
* give a tag ` Name = EFS`
* serach for EFS and and `create a file system` provide a name `tools` and slect
the VPC
* unselect the `automatic backups`
* next, select the zone from the given and and provide the security group we have
ceated
* click on create now
Note :
    while you are creating a EFS there is no need to mention the size
    because the size automatically goes.

* create two Amazon linux vms of type (Amazon linux 2)[this machines have
softwares installed] in aws and ssh through powershell
* do `sudo yum install -y amazon-efs-utils`, and for other system check the aws
docs or go to multicloud notes on 2 june you will find a link to check this
* do this on both servers
* now go to aws and open EFS , there select the on the name `tools` and select
`attach`
* when you select `attach` you will find commands to mount with server.
* copy the first command `sudo mount -t efs -o tls fs-_number_:/ efs(folder_name)`
* when you directly execute it you will get an error `mount point efs does not
exist`, because there is no folder create by name `efs`.
* to create a file do this `sudo mkdir /efs`.
* now execute the copied command again `sudo mount -t efs -o tls fs-_number_:/
/efs`
* it will mount
* do `sudo df -h`
* ls /efs
* create two files : `sudo touch /efs/1.txt` `sudo touch /efs/2.txt`
* do ls /efs/
* these files are stored in EFS, now will go to other machine and mount by
creating a new folder and try to do `ls /tools`
* you will find the files 1.txt and 2.txt
* This is how the EFS works.

* Delete everything after doing , go to EFS and delete file share manually, it
won't delete along with vms.

* Now FSx:
    * go to aws and search for FSx and open it, this also similar as EFS , if we
want to create a file system, it will give you all third party file systems, if
your company is using any kind from these you can use these.
    * Fsx is for third party storage offerings.
```

- try to calulate cost of EFS for 10 TB of data through aws calculator.

File storage on Azure

- Azure has Azure File Shares which is part of storage Account.
- Azure storage Accounts is one account to handle all storage types by Azure. Storage account provides
  - Blob Storage
  - File Shares
  - Disk Storage
  - table storage
  - queue storage
- Azure storage account support fileshares
- Azure supports two types of file shares sku
  - standard (Genral purpose account)
  - Premium (premium azure file share)
- **Practical:**

```
* Go to azure and open storge account and create a storage.
* name should be unique.
* in both standard and premium we have file share option now we will go with
stadard
* now create the storage account.
* after creating the storage account from left side options you can see that
fileshare option select it.
* create a new file share
* provide name as `tools`
* tir:
    * Cool : you don't use it frequently
    * Hot : you use it frequently
    * Transaction optmizer : work normally
*  we can see that the mximun capacity is 100TB.
* next in backup, unselect it. we will learn all types of backups in one class. we
can take all backups from one center location thts is where our focus
* now create it .
* you will be charged for used capacity in all the clouds.un like disks because
when you create a disk of 4/8 Gb it is chargable
* Now Create a Linux VM :
-------------------------
    * login to linux vm from poweshell
    * now go to storage account which is created now from there go to file share
and on top left corner you can see the `connect` button
    click  on it.
    * it will ask for machine select the machine you are using
    * Write the path where you wanted it to mount, in our case we have created
linux machine , in the box w will write as`/tools` click on csript button it will
show you script now copy the script
    * go to vm now do this`vi mounthelper.sh`
    * it wil take you inside the file
    `
    #!/bin/bash
    _paste the content_
    `
    save it `:wq`
    *  do `chmod +x monthelper.sh`
    * `./mounthelper.sh`
    * `sudo df -h`
    * `suod touch /mnt/tools/{1..100}.txt
    * `ls /mnt/tools` to check files
    * Azure one more feature
        * go to azure -> storage account created -> file shres -> form left side we can see `Browse` click on it you will be able to see the files which created in server.
* Now we will go for Windows mahine:
----------------------------------
    * go to run, mstsc -v _ip_ address provide username and password connect to vm .
    * go to file share from azure and click on connect , now select windows select some drive letter like `p`.
    * same as earlier copy the script
    * open notepad file and paste it with name and extension as `mount.ps1`
    * open your powershell navigate to the file location and execute the file.
    * it will run and shows the data
    * right click on ThisPC and map a network drive, give some different drive name as `z` and you will find url (ex: \\Itfilesharedemojune24.file.core.windows.net\\tools)in script
    * in place of folder path paste this url it will show you the files.
    * this way you can access the network fileshare
```

- Azure third party storage solutions
  - Azure Netapp
  - Azure lustre
- go to azure , in global search , search for `NetApp` you will get it.
- same search for managed cluster.

3 June :

Object Storage:

- Also reffered as blob Storage(Binary large object) which is nothing but a file
- Object Storage:
  - Store any file
  - think of as unlimited
  - Access the data using http urls.
- AWS has a service called as s3 (simple storage service) which launched object storage
  - store any file of any type (each file size cannot be grater than t TB)
  - Storage size unlimited

  - Access the data using https urls.

- Azure has launched storage account in which we have Blob storage for the same
  - Store any file of any type (each file size cannot be grater than 4.7 TB)
- This type of storage has started new way of storing
  - Google Drive/One drive/icloud
  - Online vedio and audio streaming platforms
  - Backup and archival solutiuons
  - Media on websites
- New opportunities
  - cheaper way for hosting static websites

```
what is Data engineering:

companies are having a large number of data like
excels,files,docs,images,db's ..etc to convert this data into
information(that is useful information) few years ago there is launched
Bigdata, this is collect all the data and by using Machine learning and Ai
this will be converterd into information.
* data engineergs in organisation are responsible to collect all the data
and store them in s3,blob storages as data lakes.
* using object storage as data lakes.
```

AWS S3:

- S3 organises data under buckets.
- Bucket names are unique across aws acounts.
- Buckets will have
  - pdf
  - image
  - Vedio
- Bucket belongs to a region.

Practical:

AWS S3:

```
* Go to AWS and open S3
* Create a new bucket with unique name. ex:`storage4589y986`
* when it comes to object ownership , ideally we will be using `ACL's
disabled(recommended)`, but now we will go with `ACLs enabled`.
* uncheck the bock `unblock all public access`, check the box
* without changing any options create a bucket
* Lets upload a pdf file to our bucket
* in `obkects` section you can see the upload using that upload a pdf file by drag
and drop
* once you dropped scroll below and open `permissions`under `Access control
list(ACL)` check the box `Grant public-read access` check one more box show in
```

```
below to it.
* now come back , you can see the file is uploaded, now copy the url from it.
* there is tab `copy url` click on it that will copy url and try to access over web.
* pattren : `https://<bucket-name>.s3.<region>.amazonaws.com/<object-path>`
```

Azure Blob storage:

- Azure blob storage is part of storage account and the storage for any file type is called Block Blob
- Azure support 3 types of Blobs
  - Block Blob
  - Append Blob (used for logging)
  - Page Blob (virtual hard disk)

```
* Go to Azure and open storage accounts and create one
* create a container with name `testing` and you can see there `Anonymous access level` written as private
* go to `configuration` from settings and there you will find `Allow Blob anonymous access` you need to chage it to `Allow` from `Disabled`.
* now try again to create a container with name `testing` this time you will find option is enabled select anonymous access level to ` Container level`
* crete the container.
* upload one pdf file to container and get pdf link by selecting the uploaded pdf, copy the link from properties and access it through the web browser.
```

5 June :

Repository:

- folders are used to store files and data.

- where as Repositories are stores `Versions`, whenever you need lower versions you can get it.

- S3 supports Versioning

- Create S3 bucket in AWS

- Versioning can be enabled and suspended in AWS:

  - Example:
    - In S3 bucket you have taken a file and enabled versioning and upload three versions of it , you can't disable the vesioning means that deleting the older versions, thats why we have option `suspended` not `disabled`.

- go to bucket and select `properties` from the label you can see Bucket versioning and `edit`.

- click on `edit` you will see two options `suspend` and `enable` select enable and save it.

- if you try to upload a same file with some changes, it will try to create versioning
    - Try this:
        - upload a text file which is having text as `hello` in it and reupload the same file by adding one more line to it `how are you`, after uploading them.
        - go to bucket again and find a on/off toggle button with show versions, this will show versions.
- Now , when you `suspend` versions
    - go to properties and `suspend` versioning
    - from now the earlier versions will remain same and new versions won't create from now.

S3 Bucket Pricing

- Factors that impact cost of s3 are
    - Size
    - data transfer
    - redundancy
- [Netfilx storage example] -> 10000 users watching * 1 Gb vedio = 10000 GB data transfer, this is cost effective.
- In AWS the servers are present in availability zones thes zones are presnet inside of regions.
- AWS will takes multiple copies of data and keeps in different AZones, even one Azone is down we have other zone to get data.
- Durability:
    - it defines as `what is the chance of file getting crash`
    - ex: CD's not durable there are chances of getting scratches or some crashes.
- S3 storage Classes :S3 Storage class helps in selecting
    - Cost of storage
    - Cost of data transfer
    - redundancy levels
- S3 Storage classes
    - Intellegent tier
        - when you don't know which tier you need to use, then select this . there a machine learning model that runs on your data patterns and cheks the activity and select the tier automatically.
    - standard [in AWS free tier you will get 5gb of standard storage free]
        - high storage cost
        - low data transfer cost
    - standard -ia(in frequent access):
        - low storage cost
        - high data transfer cost
    - one Zone -ia
        - low storage cost
        - high data transfer cost
        - reduced redundancy
    - Glacier
        - no data transfer (not possible)

        - storage cost is cheapest(chepest form of storge)
    - Reduced redundancy
        - high storage cost
        - low data transfer cost

Watch class vedio from 47:00 after for checking calculations.

6 June

Azure Storage account - Block Blob Storage

- Three factors:

    - Size
    - Data transfer
    - Redundancy

- Redundancy (3 copies):

    - LRS (Replicates data inside same data center three times)
    - ZRS (Replicates data each copy in different data centers in a same region)
    - GRS (Replicates data in two different regions following LRS in each region)
    - GZRS (Replicates data in two different regions follwing ZRS in primary and LRS in secondary region)
    - RA-GRS (Read-Access Geo-Redundant Storage): Provides geo-redundant storage with read access to the secondary region, ensuring high availability and disaster recovery.
    - GRS (Geo-Redundant Storage): Replicates data to a secondary region for disaster recovery but does not provide read access to the replicated data in the secondary region.

- Tiers:

    - Hot
    - Cool
    - Cold
    - Archive

```
Hot Tier:

Purpose: For data that is accessed frequently.
Access Cost: Lowest cost for accessing data.
Storage Cost: Higher storage cost compared to other tiers.
Minimum Retention Period: No minimum retention period.
Typical Use Cases: Frequently accessed files, active documents, daily backups.

Cool Tier:

Purpose: For data that is infrequently accessed and stored for at least 30 days.
Access Cost: Higher cost for accessing data compared to the Hot tier.
Storage Cost: Lower storage cost compared to the Hot tier.
Minimum Retention Period: 30 days.
```

```
Typical Use Cases: Short-term backups, less frequently accessed data, disaster recovery files.

Archive Tier:

Purpose: For data that is rarely accessed and stored for at least 180 days.
Access Cost: Highest cost for accessing data.
Storage Cost: Lowest storage cost.
Minimum Retention Period: 180 days.
Typical Use Cases: Long-term backups, archival storage, compliance and legal data.
These tiers help manage costs effectively by aligning the cost structure with data access patterns.
```

```
Example: Backup Data Storage

Scenario:
You are an IT administrator managing backup data for your company. You perform weekly backups of important files and want to store these backups in a cost-effective manner.

Step-by-Step Process:

Initial Storage:

You perform a backup on July 1st and store the backup data in the Cool tier.
Azure requires that this data remains in the Cool tier for at least 30 days.

Minimum Retention Period:

From July 1st to July 31st, the data is in the Cool tier.
During this 30-day period, you cannot move this data to another tier (like Hot or Archive) or delete it without incurring early deletion charges.

Post-Retention Period:

After July 31st, the data has met the minimum retention period requirement.

Starting from August 1st, you can choose to either:

Keep the data in the Cool tier if you expect it will be accessed infrequently and you want to continue saving on storage costs.
Move the data to the Archive tier for long-term storage at an even lower cost if you do not anticipate accessing it often.
Move the data to the Hot tier if you now need frequent access to it.
Delete the data if it is no longer needed.

Continued Storage:

If you decide to keep the backup data in the Cool tier, it will remain there until you take action.
You will not be charged any early deletion fees since the minimum retention period
```

```
has already passed.

Summary:

July 1st: Backup data is stored in the Cool tier.
July 1st - July 31st: Minimum retention period during which the data must remain
in the Cool tier.
August 1st onwards: You have the flexibility to move, delete, or keep the data in
the Cool tier without any additional charges.
This example illustrates how the Cool tier's retention period works, ensuring
cost-effectiveness while providing flexibility after the minimum period has
passed.

if your data is in Archive, you can't access it directly. you first need to
rehydrate it and move to any other tier like cold, hot to start using.

Summary for Archive data :

Possible to Access: Yes, you can access data from the Archive tier, but it
requires rehydration.
Time Required: Rehydration can take from a few hours (High Priority) to up to 15
hours (Standard Priority).
Costs Involved: Additional costs are incurred for rehydration and storing the
rehydrated data in Hot or Cool tiers.
```

- Access Tiers:
  - Hot (frequently accessig data)
  - Cool(infrequently accessing data)[30 days retention period]
  - Cold(Rarely accessing data)[90 days retention period]
  - Archive(you don't access this data, since it is called offline tier)
- In AWS we call Storage class, in Azure we call it as Access tier
- In AWS when you are creating a S3 bucket you set storage class, when you are setting storage class you are setting two things `what is redundancy level & what is access patterns`
- In case of Azure you have to set two different things `redundancy level & Access tier` **watch the class vedio from 9:00 to 16:00[pricing calculations]**

Practical

- Go to Azure create a Storage account.

- you can see the tires mentioned there

  - Premium (this is faster)
  - Standard

- As of now we will go with Standard

- on storage overview page if check for Replication you will find the redundancy type

- create a new container, try to upload file in Advanced you will find tiers select Hot and upload it.

- now upload the same file even if it shows file already exists.

- now open the file and go to the Versions tab `you won't find any versions`
- you need to set versions:
  - with in the storage account from left side options select `Data protection` and scroll down you will find a check box `Enable Versioning for Blobs` expand it from there select `Keep all versions` and save it.
- now again upload the same file with some change in data, check for versions you will find the versions.
- Lifecycle transitions: This is about changing the access tier of an object according to time periods.
- from left side options select `life cycle transactions` be in the list view and `add new rule`
  - Rule name : Movement
  - Apply rule to all the blobs in storage account
  - Limit blobs with filters (if you want to this for particular files like all .mp4 files[*.mp4] we can use this option)
  - rest keep same and click on next
  - under `if` condition let it be Last modified and in days section give 30 days, in Then condition give `Move to cool storage`
  - add one more condition ,there in days give 90 and tier is move to cold storage
  - add one more condition with 180 days for Archive
  - add one more condition with 360 days for Delete the blob.
  - [above conditions states that, when no one is accessed for a month 30 days move it to cool else if not touched for 90 days move it to cold, else if not touched for 180 move it for Archive else not touched for year 365 days delete the blobs]
  - when you are moving from one tier to another it inccur costs, same for AWS and Azure, doing too many movements is also not good.
- IN AWS if i want data to be copied in other region :
  - go to AWS S3 buckets:
  - open any storage account if you have else create one and open it inside you will find management section , inside you will find Replication rules.
  - create a replication rule
    - provide a name for rule
    - enable `apply to all objects`
    - choose the buckets in this account
    - create one more bucket in different region,enable bucket versioning
    - go to replication rule , provide name
    - Apply all the bukets
    - specify the buckent in your account
    - choose the previous bucket, click on choose path and select create IAM rule, dont change the storage class click on save
    - yes , replicate objects
    - provide path if you got error
- In both Azure and AWS we have replication rules the purpose of replicatin rule is when you want to share the contents of buckets to others , the other could be same region a bucket of yours or bucket of

others and that bucket can be present in same account or in a different account

- Object Replication in Azure
    - you have a bucket contains audio, vedio, image files and you want to sync only images to another storage account you need use object replication
- replication are done externally by users, and the redundancy are done internally
- Scenario: replication rules are generelly used for sharing data, you have financial statements that are in a bucket which is private to you, now some other company like delloitte is doing audit on your financial statements so, you need to share the data with them so you need to select data transfer for another account and later you need to write replication rule

Hosting Static websites

- Static websites use html, css and javascript **try to host a web site in linux machine**
- now go to AWS create a new bucket upload only files to bucket
- now go to properties and select static web site hosting, there in index document write `index.html` and save changes.
- it will give you link when you come out of it.
- make sure that all the files are public
- select all the files and click on actions and make it as public. now try to access it from website, it should work.
- go to latency testing website and try to do test from different regions and observe the latency timing

7 June

Content Delivery Networks

- Let us asssume that i have server in hyd and i have uploaded a vedio
- there are twi persons watching vedio one from USA and other from Hyderabad.
- the person who is accessing vedio from USA will have performance issue because the network packets needs to travel from hyd to Usa.
- and for the person who is accessing from Hyderabad will have less buffering because distance is less. **How to solve this problem:**
    - we can solve this by means of using content delivery networks
    - apart from the original location, through out the world they will other locations .
    - the vedio which is present in the server will be copied to all the nearby locations called as edge locations through out the world.
    - now whenever USA person try to watch vedio this will be playing from the near by location this idea reduces the latency and give better performance , the guys who access from hyd and USA will have similar performance.
    - there is term called `Time to Live` , which means the time period of a vedio that stays in edge locations. all the files will be copied an stored in all edge locations throught the world, there is

time period for them. if no one is accessing the vedio from any one of the location it will be removed from there.

- In Aws, CDN solution called as `Cloud front` which caches the content multiple edge locations

Practical

- write a html page to view the vedios, take w3 school ref or take html file from telegram provided by sir.
- create a new bucket and disable ACL's and upload vedio files and make them as public visible.
- and copy theirs URl's put them in html file and try to open and observe the latency
- now we will enable caching in edge locations
  - search for `cloud front` open it and create a distribution
  - select the origion domain where our vedios are present
  - path : /
  - make orign : public
  - scroll down fully and navigate to settings check whether the `use all edge locations(best performance)` is enabled or not
  - dont want security protection
  - save and create.
  - copy the distribution name and replace this in html file , from now the link will be changed
  - wait for the deployment fishes.
  - and try to access the links after deploymet, by this way you can improve performance
- create one more distribution, this time use one of the website from free css.com
- provide origin domain and if any notification with `Use website endpoint` click on it.
- origin path : `/`
- select `dont enable security protection`
- save and create it
- copy the domain name and check whether it is working or not from browser
- now test the latency by using website `website speed test`
- this is how cloud front helps to caching websites and make data available in all edge locations.
- we can restrict users from some geographies by providing access to limited locations

Azure CDN

- go to azure and create a storage account , inside of it create a container and provide anonymous read access to it
- upload the same vedios here in cotainer

- azure CDN also have multiple locations around the world , azure also has some other tieups with third parties
- now in global search `azure front door cdn profiles`
- create new, select azure front door and quick create with defult settings
- in basics provide resource group name and next configuration
- in configuration with in step one you will observe `frontends/domains` in fornt of it there will be `+` icon select it and provide details
    - provide host name `ltvedios`
    - then now select `Backend pools`
        - Backend host type - select Storage(our vedios are there in storage)
        - select the storage domain
        - click on add
    - provide name if not provided
    - add routing rule
        - provide name as `all`
    - save and create it
- go to resorse and copy the url and browse you should get vedios you can change to different vedio by adding /one.mp4

11 June

Azure Storage Account

- Tiers:
    - Standard
        - offers lowest cost per gb and is backed by HDDs
    - Premium:
        - Offers low lantency performance & is backed by SSD
- Storage Account type
    - General purpose storage:
        - This is of tier standard
        - Replica options (redundancy)
            - LRS
            - ZRS
            - GRS
            - GZRS
    - Block Blob Storage
        - Redundancy options
            - LRS
            - ZRS
    - File Storage
        - Redundancy options
            - LRS

        - ZRS
  - Page Blob storage
    - Redundancy options
      - LRS *ZRS

Storage Account Endpoints

- The name for storage account must be unique.

- securing the endpoints

  - open `networking` which is under `Security+networking`
  - if you see there it is enabled for all the networks `enabled from all networks`
  - i can make it available for only selected ip networks by choosing option `Enabled from selected virtual networks and ip address` and scroll down under `Firewall` you can specify the ip address
  - try to provide your ip address and verify with another devices are they able to open it.
  - even the vedio is public we can restrict the access by endpoints

- Scenario:

  - you have hosted vedios in azure storage and made them available for everyone
  - is it safe ?
  - is the data encrypted ?
  - is it secure ?

Azure Storage Security:

- Security is a major concern for organisations storing data in public cloud
- Azure provides a set of comprehensive security capabilities to address the concerns of orgaizations
  - Encryption: Azure encrypts all data written in azure storage account automatically usingstorage encryption serivce.
  - Authentication: Azure storage supports Azure AD based authentication for Azure blobs. UsingRBAC we can control the access of storage account
  - Data in transit: (HTTPS)
  - Azure Disk encryption [you can encrypt disks]
  - Shared Access signatures: We can control fine-grained access to data objects using SAS keys

Shared access signatire

- Every storage account has two keys called the access keys

- We can share these access keys with users and they can use this is Autorization header of API calls toget authenticated.

- This is not considered to be a good apporach as the user gets full access as long as keys are not rotated

- SAS is a UR that is composed of various parameters by which you can restrict access to Azure storage

- SAS gives us the ability to grant granular access to the objects. Using SAS

  - Control access at the service level
  - Set the time frame during which SAS is valid
  - Set permisions like read, write, delete etc
  - Set IP rangres from which SAS Keys are accepted
  - Set the protocol to HTTP or HTTPSCreate a SAS for read access

- practical:

  - go to azure create a storage account and create a container don't make it available for public leave as private
  - upload some images
  - now form left side options you can see `Shared access signatures` click on it go inside edit accordingly
  - using these keys you can make it available for a desired time by mentioning time and even for days.
  - you can set access level as read, write..etc
  - you can choose protocols
  - you can allow this for partical groups and people by mentioned IP address.
  - later providing everything ,click on generate SAS tocken and URL , it shows urls there copy the URL and access it by including `/image.jpg` since there are so many images in container.
  - you do this by container level , storage level too.

Data protection in azure storage accounts

- create a new resource group
- create a new storage account, while creating stop at the `Data protection tab`
- observe all the things mentioned over there as soft delete, enable versioning .
- save and create it
- create a container and upload some items to it.
- Since soft is enabled , after deleting within 7 days we can revert them again
- delete any one of the item from them , after deleting you can see there is an option `show delete files` enable it.
- you can see all the deleted files, select the deleted file options button and select undelete. it will be back in your container again.
- you can udelete files with in 7 days of deleted date.
- go to resource group and delete it.
- after deleting the RG, go to storage accounts you can see option `Restore` click on it, this will show all the deleted accounts 14 days back.
- select the account and which you want to restore and for this it needs the RG with old name re-created.
- if you delete anything like container[7 days], storage account[14 days] you can recover it but there is a time frame associated with it.
- now go to `data protection` option from storage account and you can see the soft delete , you can put it for 365 days default is 7.

12 June

- Presined URLs: ref : https://docs.aws.amazon.com/AmazonS3/latest/userguide/ShareObjectPreSignedURL.html

- create a s3 bucket and upload some files with private access
- now select one of the uploaded object and click on actions you can see an option `Share with presigned URL` click on it
- it will ask you to provide time in min,provide time and select `create presigned URL`.
- at the top right corner on a green label you will see `copy presigned URL` . copy from there and browse it.
- you will be having access to it until the desired time you gave .

Recovery from deleted objects in s3

- S3 cannot recover objects if the versioning is disabled.
- If the versioning is enabled, we can use delete markers to recover the deleted objects
- upload a files by means of versioning and delete it , now follow below link to recover them
- https://repost.aws/knowledge-center/s3-undelete-configuration
- you can recover versioning objects from deletion
- now configure Aws CLI, create an IAM user and assign Admin role
- go to user profile now , you can see security credentials and create access key and go to powershell enter `aws configure`. it will ask for access keyid and secret access key enter you will be connected.
- next it will ask for region name give any region `ap-south-1`
- now enter the command to check the s3 bucket list `aws s3 ls`
- if you got the list your connection is made correctly. **delete the access key once the work is done**

Commands:

```
<executable> [args]

ping google.com

cp 1.txt 2.txt

* arguments change from tool to tool, based on developer who developed .
-> aws s3 ls
-> az storage account list

Aws:
* aws <service> <action> [arguments]
* aws s3 ls
* aws ec2 start-instance --instance-id "i-129654654654"

* you dont need to remember all of these do as below
    * go to browser type `aws s3 cli` ` aws <service> cli` this will provide you commands that supports
    * if you see inside command, anything you see in [ ] brackets are optional
```

- How AWS CLI work :
  - when you install and configure aws cli there are two files created in home directory
    - ~/aws/credentials

        - ~/aws/config
    - when you are execting cli it will pull credentials from credentials file
        - go to check c:/users/.aws/[config,credentials]

if you use python or aws cli this will take credentials from the file credentials since we call them programatic access and then it will sends the request to API this will speak with AWS cloud and creates infrastructure. Similary if you are using aws in browser, it will directly speak with API and creates infra

- python/aws cli ---> take credntls from file ---> API ---> AWS Cloud
- Browser ----> API ----> AWS cloud.

Lets create a bucket and upload files form cli

- In AWS S3 URI is a path locating your file.

- S3Uri :

    - bucket : `s3:///file.txt
    -

- it is better to save files in different folder, because whenever there are high number of users accessing same folder performance will be reduced, S3 performance will work on folder level(each folder scales unique).

- create a bucket `go to browser and serach for`aws s3 bucket cli

    - refer the commands given there
    - mb(make bucket)
    - rb(remove bucket)
    - cp(copy items)
    - ls(list objects) ..etc
    - Now go to cli do this `aws s3 mb s3://myuniquebucketname`
    - if no one created with this name a bucket will create, if it creates execute this `aws s3 ls`
    - now upload files into bucket allows
        - s3 to s3
        - local to s3
        - s3 to local
    - execute this `aws s3 cp .\storage1.png s3://bucketname` [make sure you are in right path where image(storage1.png) is present in powershell].
    - now do `aws s3 ls s3://bucetname` will list out all the files inside the bucket.
    - Now sync all the items in one folder of my system to aws s3 bucket
    - execute `aws s3 sync .s3://bucketname` this will sync all the data to bcuket.
    - To remove bucket `aws s3 rm s3://bucketname/file.txt`
    - task : find a way to delete all the files in the bucket in a single shot.

13 June

Managing AWS s3 using CLI

- Configure aws cli and keep ready by checking with command `aws s3 ls`, if you got responce for this you are good to go.
- create a new bucket with name `aws s3 mb s3://lts3demo1306.
- navigate to the folder in powershell where your files are present for uploading, and executer this to upload `aws cp .\storage1.png s3://lts3demo1306/`
- Execute this to check `aws s3 ls s3 s3://lts3demo1306/`
- Approach 2: through `aws s3api`
- create bucket `aws s3api create-bucket --acl private --bucket "s3://lts3demo1306/" --region "ap-south-1" --create--bucket-configuration LocationConstraint=ap-south-1`
- Upload a file :
- `aws s3api put-object --bucket "s3://lts3demo1306/" --key "images/storage1.png" --body .\storage1.png`

for more like this ref class vedio from 21:00

Azure cli

- configure Azure cli `az login`
- syntax
  - `az <command group> [<sub-command-group>] <action> [{--arg1 value1}...{--argn valuen}]`
- to list out all resource groups
  - `az group list`
  - ref here https://learn.microsoft.com/en-us/cli/azure/group?view=azure-cli-latest or search in browser `azure <resource name> cli`
  - `az group list --output table`
  - ref the link and try to delete th resource group
    - `az group delete --name original --yes --no-wait`

14 June

- we will try to use azure cli through visual studio code
  - go to vs code search for `azure cli tools` from extensions install it.
  - file exrension should be `.azcli`
- for icons install `vscode -icons` to get icons.
- create a file `test.azcli` inside it write commads
  - `az group create --location eastus --name storagedemo`, while you are writing this tool will help you by showing the suggestions. **ref class vedio for learning how to use cli in some more areas**

------------------------Storage END ...______________________