AWS workflow

PROCESS TO RUN PYTHON CODE ON AWS EC2 WITH DATA STORED ON S3

- Created new EC2 instance (Linux) with a 20GB EBS attached (enough disk space for data from S3)
- Saved keypair .pem file on local machine (aws_20181230.pem)
- Went into .ssh folder and ran "chmod 400 aws_20181230.pem" to set read-only permissions on the file
- Mounted EBS volume on EC2 instance per instructions on https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-using-volumes.html
  - connected to instance using SSH (ssh -i ~/.ssh/aws_20181230.pem ec2-user@ec2-52-202-215-134.compute-1.amazonaws.com)
  - ran "lsblk" command to view available disk drives
  - created file system on the device
    - ran "sudo file -s /dev/xvd..." to check if there is a file system
    - ran "sudo mkfs -t ext4 /dev/xvd..." to create an ext4 file system on the volume
  - created a mount point directory for the volume
    - sudo mkdir s3data
  - mounted the volume at the location above
    - sudo mount /dev/xvd... s3data
  - set appropriate permissions on the volume mount so that I can write to it
    - chmod 777 /s3data (per https://stackoverflow.com/questions/25838962/permission-denied-when-accessing-new-ebs-volume & https://stackoverflow.com/questions/20939562/scp-permission-denied-publickey-on-ec2-only-when-using-r-flag-on-directories)
      - SECOND TIME THROUGH THIS: used "sudo chmod 777 s3data"
- Installed python3 on the instance
  - per http://outofmyhead.olssonandjones.com/2018/02/24/how-to-install-python-3-x-on-amazon-ec2-instance
    - python —version
    - sudo yum update
    - checked which Python3 packages are installed:
      - sudo yum list | grep python3
    - installed python3: sudo yum install python36 -y
    - python3 —version & which python3.6
- Created virtual environment
  - per https://aws.amazon.com/premiumsupport/knowledge-center/python-boto3-virtualenv
    - under /home/ec2-user directory:

        - mkdir venv
        - cd venv
        - virtualenv -p /usr/bin/python3.6 python36
        - activated the environment: source /home/ec2-user/venv/python36/bin/activate
        - which python3 (to check that venv python is now the default)
- Installed OpenCV (inside the venv folder)
    - pip install opencv-python
    - checked that it works:
        - python3
        - import cv2 #no error
        - exit()
- Copied data file from S3 to EC2
    - ran aws configure on EC2
    - per https://n2ws.com/blog/how-to-guides/how-to-copy-data-from-s3-to-ebs
        - aws s3 cp s3://road-trips/DC_to_Seattle.mp4 s3data/
            - RECEIVED AN ERROR ABOUT NO SPACE LEFT ON DEVICE
                - my problem was actually that I was not in the /home directory when running the command above, but for reference, the below commands increase volume size:
                - per https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-modify-volume.html:
                    - aws ec2 modify-volume --region us-east-1 --volume-id vol-0562d5e2f036e1bfb --size 30
                    - df -h
                    - sudo resize2fs /dev/xvd...
        - checked that data was copied: ls -l s3data/
- Copied Video_Processing.py file from local machine to EC2 instance:
    - scp -i ~/.ssh/aws_20181230.pem Video_Processing.py ec2-user@ec2-52-202-215-134.compute-1.amazonaws.com:/home/ec2-user
- Ran code (inside virtual environment): python3 ....
    - Used Linux Screen (per https://linuxize.com/post/how-to-use-linux-screen and https://stackoverflow.com/questions/37796392/ec2-ssh-broken-pipe-terminates-running-process?rq=1) so that Python script does not terminate when AWS SSH connection times out
        - “screen” command
        - python3 ....py (run the desired program) (can check on the status from a different Terminal window)
        - Ctrl-a followed by Ctrl-d to detach from the screen session
- Copied output from Python script back to the S3 bucket (**39,545 files, total size: 11.9GB**)
    - From directory containing output on EC2 instance, ran: aws s3 sync . s3://

road-trips/frames_1ps
        - Additional commands for managing objects can found at: https://docs.aws.amazon.com/cli/latest/userguide/cli-services-s3-commands.html#using-s3-commands-managing-objects
- TEST RUN ONLY: Copied program output (folder of JPG files) back to local machine:
    - scp -r -i ~/.ssh/aws_20181230.pem ec2-user@ec2-52-202-215-134.compute-1.amazonaws.com:/home/ec2-user/frames_lmtd frames_lmtd
        - frames_lmtd folder does not need to exist on local machine
- Deactivated virtual environment
    - deactivate
- Exited / logged out of EC2 instance:
    - disown (more info at: https://unix.stackexchange.com/questions/479/keep-ssh-sessions-running-after-disconnection
    - exit (just logs out / closes the connection)