Task 7.1

7.1: Elasticity Task

Integrating LVM with Hadoop and providing Elasticity to DataNode Storage

```
https://aws.amazon.com/amazon-linux-2/
25 package(s) needed for security, out of 39 available
Run "sudo yum update" to apply all updates.
[ec2-user@ip-172-31-7-110 ~]$ sudo su - root
Last login: Fri Oct 30 07:24:37 UTC 2020 on pts/0
[root@ip-172-31-7-110 ~]# fdisk -l
Disk /dev/xvda: 8 GiB, 8589934592 bytes, 16777216 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: gpt
Disk identifier: 66B3909F-969E-4FD1-901C-CEE3A9974A83

Device       Start      End  Sectors Size Type
/dev/xvda1    4096 16777182 16773087   8G Linux filesystem
/dev/xvda128  2048     4095     2048   1M BIOS boot

Partition table entries are not in disk order.

Disk /dev/xvdf: 20 GiB, 21474836480 bytes, 41943040 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk /dev/xvdg: 30 GiB, 32212254720 bytes, 62914560 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
[root@ip-172-31-7-110 ~]#
```

Attached two separate EBS storage(Hard disks) of size 20 GB and 30 GB with device names as
/dev/xvdf - HD1
/dev/xvdg - HD2

```
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
[root@ip-172-31-7-110 ~]# pvcreate /dev/xvdf
  Physical volume "/dev/xvdf" successfully created.
[root@ip-172-31-7-110 ~]# pvcreate /dev/xvdg
  Physical volume "/dev/xvdg" successfully created.
[root@ip-172-31-7-110 ~]# pvdisplay
  "/dev/sdf" is a new physical volume of "20.00 GiB"
  --- NEW Physical volume ---
  PV Name               /dev/sdf
  VG Name
  PV Size               20.00 GiB
  Allocatable           NO
  PE Size               0
  Total PE              0
  Free PE               0
  Allocated PE          0
  PV UUID               6tx4cx-FWNw-5FxZ-zpJi-X7EI-C4t7-2HWff9

  "/dev/sdg" is a new physical volume of "30.00 GiB"
  --- NEW Physical volume ---
  PV Name               /dev/sdg
  VG Name
  PV Size               30.00 GiB
  Allocatable           NO
  PE Size               0
  Total PE              0
  Free PE               0
  Allocated PE          0
  PV UUID               UL0CFM-IdXv-omXd-uNqV-UjtX-75mj-znDCTN

[root@ip-172-31-7-110 ~]# D
```

Converted each storage(hard disks) into PV because VG gets space only from PV

Not allocated to any VG yet.

```
[root@ip-172-31-7-110 ~]# pvdisplay /dev/sdf
  "/dev/sdf" is a new physical volume of "20.00 GiB"
  --- NEW Physical volume ---
  PV Name               /dev/sdf
  VG Name
  PV Size               20.00 GiB
  Allocatable           NO
  PE Size               0
  Total PE              0
  Free PE               0
  Allocated PE          0
  PV UUID               6tx4cx-FWNw-5FxZ-zpJi-X7EI-C4t7-2HWff9

[root@ip-172-31-7-110 ~]# pvdisplay /dev/sdg
  "/dev/sdg" is a new physical volume of "30.00 GiB"
  --- NEW Physical volume ---
  PV Name               /dev/sdg
  VG Name
  PV Size               30.00 GiB
  Allocatable           NO
  PE Size               0
  Total PE              0
  Free PE               0
  Allocated PE          0
  PV UUID               UL0CFM-IdXv-omXd-uNqV-UjtX-75mj-znDCTN

[root@ip-172-31-7-110 ~]#
```

Showing info of PV1

Showing info of PV2

```
[root@ip-172-31-7-110 ~]# vgdisplay
[root@ip-172-31-7-110 ~]# vgdisplay iiecvg
  Volume group "iiecvg" not found
  Cannot process volume group iiecvg
[root@ip-172-31-7-110 ~]# vgcreate iiecvg /dev/sdf /dev/sdg
  Volume group "iiecvg" successfully created
[root@ip-172-31-7-110 ~]# vgdisplay iiecvg
  --- Volume group ---
  VG Name               iiecvg
  System ID
  Format                lvm2
  Metadata Areas        2
  Metadata Sequence No  1
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                0
  Open LV               0
  Max PV                0
  Cur PV                2
  Act PV                2
  VG Size               49.99 GiB
  PE Size               4.00 MiB
  Total PE              12798
  Alloc PE / Size       0 / 0
  Free  PE / Size       12798 / 49.99 GiB
  VG UUID               uhk8GW-sacZ-VTyw-BrjU-LtrT-rMSP-CdSjGp

[root@ip-172-31-7-110 ~]#
```

Showing all available VGs

Checking if any VG with this name 'iiecvg' available

Creating a VG with combined storage of PV1 and PV2

Info related to VG

Space available for storing is not allocated yet. First create partition to store.

Total storage 20GB + 30GB nearly 50GB is unallocated

```
[root@ip-172-31-7-110 ~]# lvcreate --size 40G --name lv1 iiecvg
  Logical volume "lv1" created.
[root@ip-172-31-7-110 ~]# vgdisplay iiecvg
  --- Volume group ---
  VG Name               iiecvg
  System ID
  Format                lvm2
  Metadata Areas        2
  Metadata Sequence No  2
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                1
  Open LV               0
  Max PV                0
  Cur PV                2
  Act PV                2
  VG Size               49.99 GiB
  PE Size               4.00 MiB
  Total PE              12798
  Alloc PE / Size       10240 / 40.00 GiB
  Free  PE / Size       2558 / 9.99 GiB
  VG UUID               uhk8GW-sacZ-VTyw-BrjU-LtrT-rMSP-CdSjGp

[root@ip-172-31-7-110 ~]#
```

Creating a dynamic partition called LV(Logival Volume) inside VG of size 40GB getting storage from both PVs.

Total allocated size i.e. space available to store is 40 GB(partition size)

Unallocated Space in VG

```
[root@ip-172-31-7-110 ~]# lvdisplay iiecvg/lv1
  --- Logical volume ---
  LV Path                /dev/iiecvg/lv1
  LV Name                lv1
  VG Name                iiecvg
  LV UUID                VITKJc-QSuA-0Q2W-YJP2-zPKI-uq2e-UDKA9F
  LV Write Access        read/write
  LV Creation host, time ip-172-31-7-110.us-east-2.compute.internal, 2020-10-30 07:56:11 +0000
  LV Status              available
  # open                 0
  LV Size                40.00 GiB
  Current LE             10240
  Segments               2
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     256
  Block device           253:0
```

Info related to LV

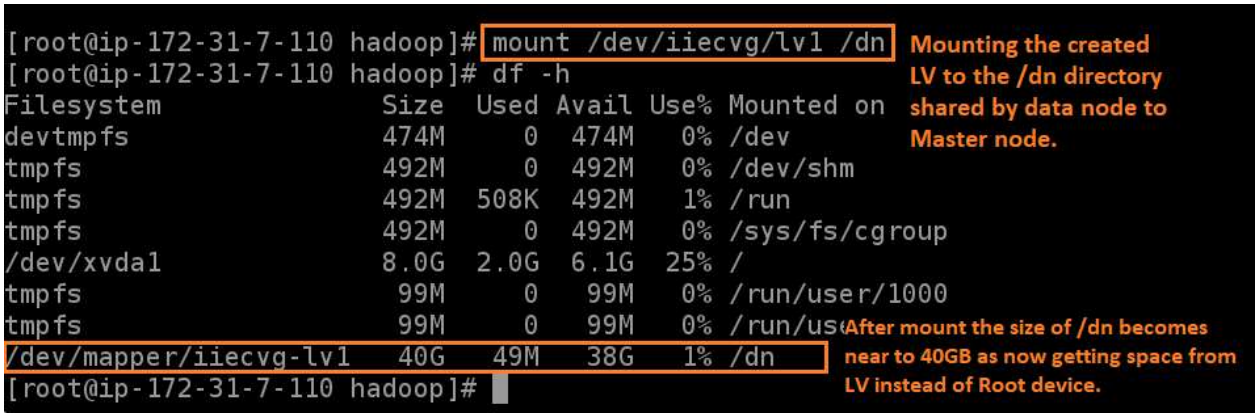
```
[root@ip-172-31-7-110 ~]# lvdisplay
  --- Logical volume ---
  LV Path                /dev/iiecvg/lv1
  LV Name                lv1
  VG Name                iiecvg
  LV UUID                VITKJc-QSuA-0Q2W-YJP2-zPKI-uq2e-UDKA9F
  LV Write Access        read/write
  LV Creation host, time ip-172-31-7-110.us-east-2.compute.internal, 2020-10-30 07:56:11 +0000
  LV Status              available
  # open                 0
  LV Size                40.00 GiB
  Current LE             10240
  Segments               2
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     256
  Block device           253:0
```

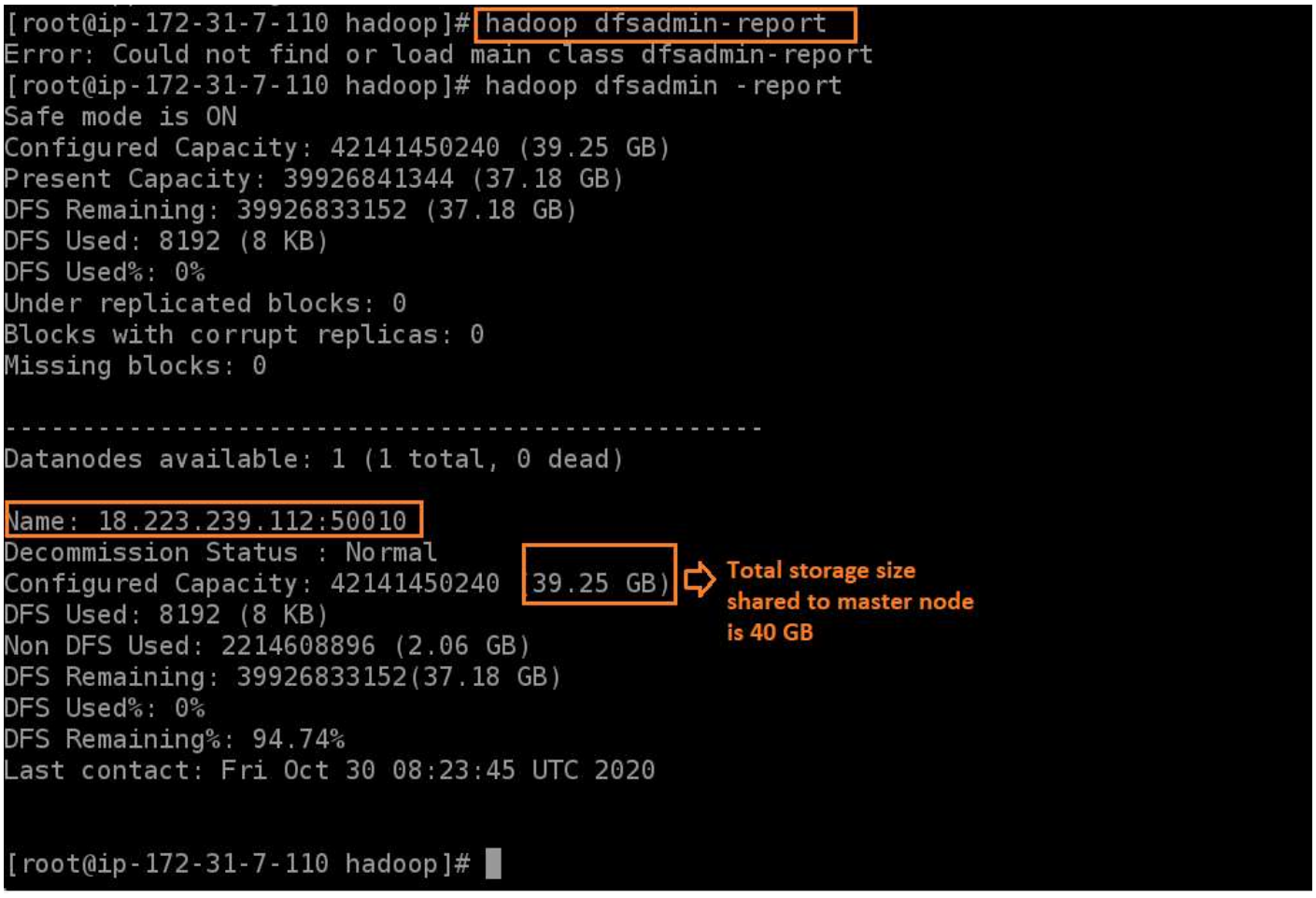
```
[root@ip-172-31-7-110 ~]# mkfs.ext4 /dev/iiecvg/lv1
mke2fs 1.42.9 (28-Dec-2013)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
2621440 inodes, 10485760 blocks
524288 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=2157969408
320 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
        4096000, 7962624

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

Formatting the LV

```
[root@ip-172-31-7-110 hadoop]# hadoop dfsadmin -report
Safe mode is ON
Configured Capacity: 8577331200 (7.99 GB)
Present Capacity: 6481682432 (6.04 GB)
DFS Remaining: 6481674240 (6.04 GB)
DFS Used: 8192 (8 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

-------------------------------------------------
Datanodes available: 1 (1 total, 0 dead)

Name: 18.223.239.112:50010
Decommission Status : Normal
Configured Capacity: 8577331200 (7.99 GB)
DFS Used: 8192 (8 KB)
Non DFS Used: 2095648768 (1.95 GB)
DFS Remaining: 6481674240(6.04 GB)
DFS Used%: 0%
DFS Remaining%: 75.57%
Last contact: Fri Oct 30 08:20:57 UTC 2020


[root@ip-172-31-7-110 hadoop]#
```

Initially the data node was sharing the storage of its Root or Slash drive via /dn directory nearly equal to 8 GB.


[root@ip-172-31-7-110 hadoop]# mount /dev/iiecvg/lv1 /dn
[root@ip-172-31-7-110 hadoop]# df -h
Filesystem Size Used Avail Use% Mounted on
devtmpfs 474M 0 474M 0% /dev
tmpfs 492M 0 492M 0% /dev/shm
tmpfs 492M 508K 492M 1% /run
tmpfs 492M 0 492M 0% /sys/fs/cgroup
/dev/xvda1 8.0G 2.0G 6.1G 25% /
tmpfs 99M 0 99M 0% /run/user/1000
tmpfs 99M 0 99M 0% /run/us
/dev/mapper/iiecvg-lv1 40G 49M 38G 1% /dn
[root@ip-172-31-7-110 hadoop]#
Mounting the created LV to the /dn directory shared by data node to Master node.
After mount the size of /dn becomes near to 40GB as now getting space from LV instead of Root device.


[root@ip-172-31-7-110 hadoop]# hadoop dfsadmin-report
Error: Could not find or load main class dfsadmin-report
[root@ip-172-31-7-110 hadoop]# hadoop dfsadmin -report
Safe mode is ON
Configured Capacity: 42141450240 (39.25 GB)
Present Capacity: 39926841344 (37.18 GB)
DFS Remaining: 39926833152 (37.18 GB)
DFS Used: 8192 (8 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

Datanodes available: 1 (1 total, 0 dead)
Name: 18.223.239.112:50010
Decommission Status : Normal
Configured Capacity: 42141450240 (39.25 GB)
DFS Used: 8192 (8 KB)
Non DFS Used: 2214608896 (2.06 GB)
DFS Remaining: 39926833152(37.18 GB)
DFS Used%: 0%
DFS Remaining%: 94.74%
Last contact: Fri Oct 30 08:23:45 UTC 2020
[root@ip-172-31-7-110 hadoop]#
Total storage size shared to master node is 40 GB

```
[root@ip-172-31-7-110 hadoop]# lvdisplay
  --- Logical volume ---
  LV Path                /dev/iiecvg/lv1
  LV Name                lv1
  VG Name                iiecvg
  LV UUID                VITKJc-QSuA-0Q2W-YJP2-zPKI-uq2e-UDKA9F
  LV Write Access        read/write
  LV Creation host, time ip-172-31-7-110.us-east-2.compute.internal, 2020-10-30 07:56:11 +0000
  LV Status              available
  # open                 1
  LV Size                40.00 GiB
  Current LE             10240
  Segments               2
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     256
  Block device           253:0

[root@ip-172-31-7-110 hadoop]# df -h
Filesystem              Size  Used Avail Use% Mounted on
devtmpfs                474M     0  474M   0% /dev
tmpfs                   492M     0  492M   0% /dev/shm
tmpfs                   492M  508K  492M   1% /run
tmpfs                   492M     0  492M   0% /sys/fs/cgroup
/dev/xvda1              8.0G  2.0G  6.1G  25% /
tmpfs                    99M     0   99M   0% /run/user/1000
tmpfs                    99M     0   99M   0% /run/user/0
/dev/mapper/iiecvg-lv1   40G   49M   38G   1% /dn
[root@ip-172-31-7-110 hadoop]#
```

#df -h command shows the total available size capable to store data(formatted space). Whereas #lvdisplay command shows the total size.

```
[root@ip-172-31-7-110 hadoop]# lvextend --size +5G /dev/iiecvg/lv1
  Size of logical volume iiecvg/lv1 changed from 40.00 GiB (10240 extents) to 45.00 GiB (11520 extents).
  Logical volume iiecvg/lv1 successfully resized.
[root@ip-172-31-7-110 hadoop]# lvdisplay /dev/iiecvg/lv1
  --- Logical volume ---
  LV Path                /dev/iiecvg/lv1
  LV Name                lv1
  VG Name                iiecvg
  LV UUID                VITKJc-QSuA-0Q2W-YJP2-zPKI-uq2e-UDKA9F
  LV Write Access        read/write
  LV Creation host, time ip-172-31-7-110.us-east-2.compute.internal, 2020-10-30 07:56:11 +0000
  LV Status              available
  # open                 1
  LV Size                45.00 GiB
  Current LE             11520
  Segments               2
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     256
  Block device           253:0

[root@ip-172-31-7-110 hadoop]# df -h
Filesystem              Size  Used Avail Use% Mounted on
devtmpfs                474M     0  474M   0% /dev
tmpfs                   492M     0  492M   0% /dev/shm
tmpfs                   492M  508K  492M   1% /run
tmpfs                   492M     0  492M   0% /sys/fs/cgroup
/dev/xvda1              8.0G  2.0G  6.1G  25% /
tmpfs                    99M     0   99M   0% /run/user/1000
tmpfs                    99M     0   99M   0% /run/user/0
/dev/mapper/iiecvg-lv1   40G   49M   38G   1% /dn
[root@ip-172-31-7-110 hadoop]#
```

Here showing the total size of LV as 45GB after extension.

Demonstrating the elasticity feature of the LV shared via /dn directory to Hadoop master. We can extend the size of the LV.

But here the extension is not reflected and available size is still nearly 40GB. This because extended part of LV is not formatted, so incapable of storing data.

```
[root@ip-172-31-7-110 hadoop]# resize2fs /dev/iiecvg/lv1
resize2fs 1.42.9 (28-Dec-2013)
Filesystem at /dev/iiecvg/lv1 is mounted on /dn; on-line resi
old_desc_blocks = 5, new_desc_blocks = 6
The filesystem on /dev/iiecvg/lv1 is now 11796480 blocks long

[root@ip-172-31-7-110 hadoop]# df -h
Filesystem              Size  Used Avail Use% Mounted on
devtmpfs                474M     0  474M   0% /dev
tmpfs                   492M     0  492M   0% /dev/shm
tmpfs                   492M  508K  492M   1% /run
tmpfs                   492M     0  492M   0% /sys/fs/cgroup
/dev/xvda1              8.0G  2.0G  6.1G  25% /
tmpfs                    99M     0   99M   0% /run/user/1000
tmpfs                    99M     0   99M   0% /run/user/0
/dev/mapper/iiecvg-lv1   45G   52M   42G   1% /dn
[root@ip-172-31-7-110 hadoop]#
```

Foramtting only the extended part without data removal from previously formatted part of LV.

Now available size extended after formatting.

```
[root@ip-172-31-7-110 hadoop]# hadoop dfsadmin -report
Safe mode is ON
Configured Capacity: 47425867776 (44.17 GB)
Present Capacity: 44970913792 (41.88 GB)
DFS Remaining: 44970893312 (41.88 GB)
DFS Used: 20480 (20 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

-------------------------------------------------
Datanodes available: 1 (1 total, 0 dead)

Name: 18.223.239.112:50010
Decommission Status : Normal
Configured Capacity: 47425867776 (44.17 GB)
DFS Used: 20480 (20 KB)
Non DFS Used: 2454953984 (2.29 GB)
DFS Remaining: 44970893312(41.88 GB)
DFS Used%: 0%
DFS Remaining%: 94.82%
Last contact: Fri Oct 30 08:34:45 UTC 2020


[root@ip-172-31-7-110 hadoop]#
```

After extending the size of LV, the shared storage also extended achieving the functionality of elasticity. So on integrating LVM with Hadoop, provided Elasticity to DataNode Storage