

## Exercise 05.01: Backup and Restore

In this exercise, you will accomplish the following:

- Take a snapshot of 'killr\_video' keyspace
- Truncate the data in 'killr\_video'
- Restore the 'killr\_video' keyspace using `sstableloader`

### Step 1: Take a Snapshot

1. Confirm there are still three (3) nodes operational and connected to the cluster; DSE-node1 is connected to the seed node in dc1. The node DSE-node3 is now connected to the first node in dc1. The node DSE-node2 is connected to its own cluster in dc2. At the end of the previous exercise, all nodes in both datacenters should be up and running. Verify this is the case. In the DSE-node1 window, run the following command:

```
nodetool status
```

2. In the previous exercise, you created a record for Fred Flintstone. Verify you can access this record in cqlsh by performing the following query (use the DSE-node1 window):

```
SELECT * FROM killr_video.user_by_email WHERE  
email='FredFlintstone@gmail.com';
```

3. Take a snapshot of the data. In each of the three (3) windows (i.e., on each node instance), execute the following command:

```
nodetool snapshot killr_video
```

4. Verify that you created a snapshot by listing out the contents of the newly created *snapshots* directory:

**NOTE:** Spawn a shell (sh) and run the command in the shell to use the wildcard

```
sh -c 'ls -al /home/ubuntu/dse/data/data/killr_video/user_by_email-  
*/snapshots/*/'
```

5. Compare the contents of the *snapshot* directory to the contents of the data directory:

```
sh -c 'ls -al /home/ubuntu/dse/data/data/killr_video/user_by_email-*/'
```

6. Note that these directories contain the same files. Recall that the snapshot is really a set of hard links to the same files that are in the data directory. However, you may also notice that there are two additional directories in that parent directory; *backups/* and *snapshots/*. As this is the parent directory, it shows where the backup files are stored.

## Step 2: Truncate the Data Prior to Restore

1. This next step will simulate an error. Imagine accidentally truncating the *user\_by\_email* table by mistake. Run the following command in `cqlsh` on the DSE-node1 instance:

```
cqlsh node1
cqlsh> TRUNCATE killr_video.user_by_email;
```

2. Confirm the data is gone by running the following command:

```
SELECT * FROM killr_video.user_by_email WHERE
email='FredFlintstone@gmail.com';
```

3. The TRUNCATE command also deleted the data from the replica nodes. Fortunately, we have snapshots. Verify the data is gone from the replicas. On DSE-node1 and in each of the other two (2) windows or nodes, run the following command:

```
sh -c 'ls -al /home/ubuntu/dse/data/data/killr_video/user_by_email-*/'
```

4. Notice that the data files are not present on either of the other two (2) nodes.
5. Notice also that truncating the table causes a snapshot to occur if `auto_snapshot:` is set to true in the *cassandra.yaml* file. In each window, this can be confirmed by executing the following command:

```
sh -c 'ls -al /home/ubuntu/dse/data/data/killr_video/user_by_email-
*/snapshots/'
```

6. Notice there are two (2) subdirectories. The subdirectory named *truncated-<SNAPSHOT ID>-user\_by\_email* is the snapshot the TRUNCATE command created. It turns out that this snapshot is identical to the one we created explicitly using `nodetool snapshot`.
7. Delete the snapshot created by the TRUNCATE command so we don't get the two (2) snapshots confused. In each of the node windows, execute the following:

```
sh -c 'rm -rf /home/ubuntu/dse/data/data/killr_video/user_by_email-
*/snapshots/*-user_by_email/'
```

## Step 3: Restore the Data Using `sstableloader`

1. Restore the data using the `sstableloader` command. The `sstableloader` command expects the data to be in folders with names corresponding to the keyspace and table

within the *data* directory. Copy the snapshot files into the *data* directory. In each window run the following command:

```
sh -c 'cp /home/ubuntu/dse/data/data/killr_video/user_by_email-*/snapshots/*/* /home/ubuntu/dse/data/data/killr_video/user_by_email-*/'
```

2. We are now ready to perform the actual `sstableloader` command. In each window, execute the following command:

```
sh -c '/home/ubuntu/dse/bin/sstableloader -d `hostname -i` /home/ubuntu/dse/data/data/killr_video/user_by_email-*/'
```

**NOTE:** The shell replaces `hostname -i` with the host's IP address.

3. In the data directory, `sstableloader` has created a new set of SSTables. In any of the windows, execute the following:

```
sh -c 'ls -al /home/ubuntu/dse/data/data/killr_video/user_by_email-*/'
```

4. Notice the new set of files.
5. Query the table to make sure everything is working as expected. Execute the following CQL command:

```
SELECT * FROM killr_video.user_by_email WHERE email='FredFlintstone@gmail.com';
```

**END OF EXERCISE**