Restore Medusa Backup - Remote

Apache Medusa gives us an ability to perform a restore on a cluster different than the on which the backup was take. In this article, we will go through a restore being performed on a Test Cluster named "**PPS Cassandra restore**".

The test cluster has the exact same configuration as that of the prod cluster named PPS Cassandra cluster comprising of 4 nodes except for the different name of the cluster. This is a provision provided my Medusa to migrate data from one cluster to another running cluster.

> Note: Medusa does not touch any other keyspace or tables during a restore activity other than the one for which backup has been performed. For example: If the backup has only 1 keyspace named PPS and the target cluster has 2 keyspaces named PPS and Test-PPS respectively, medusa while performing the restore operation will only touch the PPS keypsace (perform cleanup of PPS data and restore of PPS data), Test-PPS cluster will work as it was before the cleanup activity.

Step 1 : The state of production cluster

```
Datacenter: ppsCI
==================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load        Tokens  Owns (effective)  Host ID                               Rack
UN  10.12.77.248  2.93 GiB    256     83.5%             ae9f079e-e078-4273-b28d-6a106edd7154  rack1
UN  10.12.77.249  1.58 GiB    256     74.4%             715dce46-1517-452d-b146-c6ac4d7ae75f  rack1
UN  10.12.77.251  1.96 GiB    256     69.6%             49d79a3d-d8a8-4e29-9290-30de7b951191  rack1
UN  10.12.77.250  728.56 MiB  256     72.5%             e1b6b1df-3cba-4f84-94cc-67301a2496eb  rack1


Datacenter: ppsSI
==================

Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load        Tokens  Owns (effective)  Host ID                               Rack
UN  10.162.132.6  4.05 GiB    256     100.0%            5af1e51d-28ab-4a44-9cae-cdab27d9651b  rack1
UN  10.162.132.7  4.89 GiB    256     100.0%            b503ed77-fcc4-4a81-91a2-7248e579f196  rack1
UN  10.162.132.5  5 GiB       256     100.0%            f51b9292-5ec6-4ea9-971c-ca1f3d53c1ea  rack1
```

```
root@azuppscassandra0-az-prod-ci:~# nodetool describecluster
Cluster Information:
        Name: PPS Cassandra
        Snitch: org.apache.cassandra.locator.GossipingPropertyFileSnitch
        DynamicEndPointSnitch: enabled
        Partitioner: org.apache.cassandra.dht.Murmur3Partitioner
        Schema versions:
                955ade03-92ad-3f0d-b63d-b36370b3cc25: [10.12.77.251, 10.12.77.249, 10.12.77.250, 10.162.132.6,
10.162.132.5, 10.12.77.248, 10.162.132.7]

Stats for all nodes:
        Live: 7
        Joining: 0
        Moving: 0
        Leaving: 0
        Unreachable: 0

Data Centers:
        ppsCI #Nodes: 4 #Down: 0
        ppsSI #Nodes: 3 #Down: 0

Database versions:
        4.0.14: [10.12.77.251:7000, 10.12.77.249:7000, 10.12.77.250:7000, 10.12.77.248:7000]

        4.0.11: [10.162.132.6:7000, 10.162.132.5:7000, 10.162.132.7:7000]

Keyspaces:
        system_schema -> Replication class: LocalStrategy {}
        system -> Replication class: LocalStrategy {}
        system_auth -> Replication class: NetworkTopologyStrategy {ppsCI=3, ppsSI=3}
        system_distributed -> Replication class: NetworkTopologyStrategy {ppsCI=3, ppsSI=3}
        system_traces -> Replication class: NetworkTopologyStrategy {ppsCI=3, ppsSI=3}
        pps -> Replication class: NetworkTopologyStrategy {ppsCI=3, ppsSI=3}
```

```
cassandra@cqlsh> desc KEYSPACES  ;

pps     system_auth         system_schema  system_views
system  system_distributed  system_traces  system_virtual_schema

cassandra@cqlsh> LIST USERS ;

 name        | super | datacenters
-------------+-------+-------------
   cassandra |  True |         ALL
 ppsreadonly | False |         ALL
     ppsuser |  True |         ALL

(3 rows)
```

Step 2 : Configure and create the new cluster (test-cluster in this case) similar to the one that is being restored. Setup Medusa with the same bucket name as used for backup in the original cluster

Step 3 : Run the medusa restore command and as below where you will have to mention which node data is to be restored on which node.

The generic command for restore will be as follows.

```
medusa  --prefix <prefix of the backup> --fqdn <IP address of the node of which the data is to be restored on
the node where this command is run>restore-node --remote --backup-name <backup-name> --seeds <comma separated
seed list if multiple>
```

Example : There are 4 nodes in PPS cluster and 4 nodes in PPS Cassandra Restore cluster as follows

The restore command for PPS Cassandra Restore nodes will be as follows

10.12.72.148

```
medusa --prefix pps-cassandra_WEEK_20_2025 --fqdn 10.12.77.251 restore-node --remote --backup-name backup-16-May-2025-00-00 --seeds 10.12.72.141
```

The above restore command means that on node **10.12.72.148** , the data of node **10.12.77.251** will be restored.

10.12.72.141

```
medusa  --prefix pps-cassandra_WEEK_20_2025 --fqdn 10.12.77.248 restore-node --remote --backup-name backup-16-May-2025-00-00
```

10.12.74.152

```
medusa --prefix pps-cassandra_WEEK_20_2025 --fqdn 10.12.77.249 restore-node --remote --backup-name backup-16-May-2025-00-00 --seeds 10.12.72.141
```

```
medusa --prefix pps-cassandra_WEEK_20_2025 --fqdn 10.12.77.250 restore-node --remote --backup-name backup-16-May-2025-00-00 --seeds 10.12.72.141
```

10.12.74.241

Step 4 : Logs during restore

[2025-05-16 16:35:02,093] DEBUG: Loading configuration from /etc/medusa/medusa.ini
[2025-05-16 16:35:02,097] DEBUG: Logging to file options: LoggingConfig(enabled='1', file='/var/log/medusa/medusa.log', format='[%(asctime)s] %(levelname)s: %(message)s', level='INFO', maxBytes='20000000', backupCount='50')
[2025-05-16 16:35:02,098] DEBUG: Loading storage_provider: azure_blobs
[2025-05-16 16:35:02,132] DEBUG: [Storage] Getting object 10.12.77.251/backup-16-May-2025-00-00/meta/differential
[2025-05-16 16:35:02,133] DEBUG: Using selector: GeventSelector
[2025-05-16 16:35:02,246] DEBUG: Blob pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/schema.cql was not found in cache.
[2025-05-16 16:35:02,246] DEBUG: [Storage] Getting object pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/schema.cql
[2025-05-16 16:35:02,291] DEBUG: [Storage] Getting object pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/manifest.json
[2025-05-16 16:35:02,336] DEBUG: [Storage] Reading blob pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/manifest.json...
[2025-05-16 16:35:02,412] DEBUG: This server has systemd: True
[2025-05-16 16:35:02,513] INFO: Downloading data from backup to /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03
[2025-05-16 16:35:02,514] DEBUG: Download size: 2127372256, available space: 52651438080
[2025-05-16 16:35:02,514] DEBUG: Loading storage_provider: azure_blobs

```
[2025-05-16 16:35:05,383] DEBUG: [Azure Storage] Downloading with 1 workers: cassandrafullbackups/pps-cassandra_WEEK_20_2025/10.12.77.251/data/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Statistics.db -> /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Statistics.db
[2025-05-16 16:35:05,449] DEBUG: [Azure Storage] Downloading with 1 workers: cassandrafullbackups/pps-cassandra_WEEK_20_2025/10.12.77.251/data/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Digest.crc32 -> /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Digest.crc32
[2025-05-16 16:35:05,514] DEBUG: [Azure Storage] Downloading with 1 workers: cassandrafullbackups/pps-cassandra_WEEK_20_2025/10.12.77.251/data/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Filter.db -> /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Filter.db
[2025-05-16 16:35:05,580] DEBUG: [Azure Storage] Downloading with 1 workers: cassandrafullbackups/pps-cassandra_WEEK_20_2025/10.12.77.251/data/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-CompressionInfo.db -> /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-CompressionInfo.db
[2025-05-16 16:35:05,646] DEBUG: [Azure Storage] Downloading with 1 workers: cassandrafullbackups/pps-cassandra_WEEK_20_2025/10.12.77.251/data/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-TOC.txt -> /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-TOC.txt
[2025-05-16 16:35:05,713] DEBUG: [Azure Storage] Downloading with 1 workers: cassandrafullbackups/pps-cassandra_WEEK_20_2025/10.12.77.251/data/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Index.db -> /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/local-7ad54392bcdd35a684174e047860b377/nb-8-big-Index.db
```

```
[2025-05-16 16:37:01,029] DEBUG: Disconnecting from Azure Storage
[2025-05-16 16:37:01,030] INFO: Stopping Cassandra
[2025-05-16 16:37:01,047] INFO: Waiting for Cassandra to go down on 10.12.72.148
[2025-05-16 16:37:01,047] INFO: Verifying node state for host 10.12.72.148 using check type cql
[2025-05-16 16:37:01,056] DEBUG: This server has systemd: True
[2025-05-16 16:37:01,155] DEBUG: Checking Cassandra health type: cql for host: 10.12.72.148 release_ver: 3.11.9 native_port: 9042 storage_port: 7000, rpc_port: 9160
[2025-05-16 16:37:01,156] DEBUG: Port '9042' is closed, assuming '10.12.72.148' is down.
[2025-05-16 16:37:01,156] DEBUG: The node 10.12.72.148 is not up yet...
[2025-05-16 16:37:01,156] DEBUG: Port '7000' is closed, assuming '10.12.72.148' is down.
[2025-05-16 16:37:01,156] DEBUG: The node 10.12.72.148 is not up yet...
[2025-05-16 16:37:01,156] INFO: Cassandra is down 10.12.72.148
[2025-05-16 16:37:01,156] DEBUG: Blob pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/server_version.json was not found in cache.
[2025-05-16 16:37:01,156] DEBUG: [Storage] Getting object pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/server_version.json
[2025-05-16 16:37:01,337] DEBUG: [Storage] Reading blob pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-2025-00-00/meta/server_version.json...

[2025-05-16 16:37:01,374] INFO: Moving backup data to Cassandra data directory
[2025-05-16 16:37:01,374] DEBUG: Creating directory /myntra/cassandra/data/system/repairs-a3d277d1cfaf36f5a2a738d5eea9ad6a
[2025-05-16 16:37:01,392] DEBUG: Skipping the actual restore of repairs-a3d277d1cfaf36f5a2a738d5eea9ad6a - table empty
[2025-05-16 16:37:01,392] DEBUG: Creating directory /myntra/cassandra/data/system/peers-37f71aca7dc2383ba70672528af04d4f
[2025-05-16 16:37:01,409] DEBUG: Skipping the actual restore of peers-37f71aca7dc2383ba70672528af04d4f
[2025-05-16 16:37:01,409] DEBUG: Creating directory /myntra/cassandra/data/system/IndexInfo-9f5c6374d48532299a0a5094af9ad1e3
[2025-05-16 16:37:01,424] DEBUG: Restoring /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03/system/IndexInfo-9f5c6374d48532299a0a5094af9ad1e3 -> /myntra/cassandra/data/system/IndexInfo-9f5c6374d48532299a0a5094af9ad1e3
```

```
[2025-05-16 16:37:07,725] INFO: Verifying node state for host 10.12.72.141 using check type cql
[2025-05-16 16:37:07,734] DEBUG: This server has systemd: True
[2025-05-16 16:37:07,835] DEBUG: Checking Cassandra health type: cql for host: 10.12.72.141 release_ver: 3.11.9
native_port: 9042 storage_port: 7000, rpc_port: 9160
[2025-05-16 16:37:07,836] DEBUG: Port '9042' is closed, assuming '10.12.72.141' is down.
[2025-05-16 16:37:07,837] DEBUG: The node 10.12.72.141 is not up yet...
[2025-05-16 16:37:07,837] DEBUG: Port '7000' is closed, assuming '10.12.72.141' is down.
[2025-05-16 16:37:07,837] DEBUG: The node 10.12.72.141 is not up yet...
[2025-05-16 16:37:07,837] INFO: No seeds are up yet, will wait a minute
[2025-05-16 16:38:07,837] INFO: Verifying node state for host 10.12.72.141 using check type cql
[2025-05-16 16:38:07,850] DEBUG: This server has systemd: True
[2025-05-16 16:38:07,947] DEBUG: Checking Cassandra health type: cql for host: 10.12.72.141 release_ver: 3.11.9
native_port: 9042 storage_port: 7000, rpc_port: 9160
[2025-05-16 16:38:07,948] DEBUG: Port '9042' is closed, assuming '10.12.72.141' is down.
[2025-05-16 16:38:07,948] DEBUG: The node 10.12.72.141 is not up yet...
[2025-05-16 16:38:07,949] INFO: At least one seed is now up
```

```
[2025-05-16 16:38:07,949] INFO: Starting Cassandra
[2025-05-16 16:38:07,951] DEBUG: Starting Cassandra with /etc/init.d/cassandra start
[2025-05-16 16:38:21,669] DEBUG: [Storage] Reading blob pps-cassandra_WEEK_20_2025/10.12.77.251/backup-16-May-
2025-00-00/meta/server_version.json...
[2025-05-16 16:38:21,809] DEBUG: Cleaning (/tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03)
[2025-05-16 16:38:21,810] DEBUG: Remove folder /tmp/medusa-restore-b66cd4a0-5af6-4459-aeb7-3bccb4763c03 and
content
[2025-05-16 16:38:21,845] DEBUG: Disconnecting from Azure Storage
```

Step 5 : Validate the cluster

```
root@poojatestcassandra1-az-prod-ci:~# nodetool status
Datacenter: ppsCI
==================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load       Tokens  Owns (effective)  Host ID                               Rack
UN  10.12.74.152  1.59 GiB   256     74.4%             337462b4-86b0-4fe0-976c-ad257be4d825  rack1
UN  10.12.72.148  1.96 GiB   256     69.6%             d0a74f05-34bf-408f-966f-ce3d72738f83  rack1
UN  10.12.74.241  728.56 MiB  256   72.5%             156c3db0-65a6-42ee-b8a7-afa07e72eebe  rack1
UN  10.12.72.141  2.93 GiB   256     83.5%             5bca36e5-67d7-4784-8720-a2f9d2fc0a83  rack1
```

```
root@poojatestcassandra1-az-prod-ci:~# nodetool describecluster
Cluster Information:
    Name: PPS Cassandra restore
    Snitch: org.apache.cassandra.locator.GossipingPropertyFileSnitch
    DynamicEndPointSnitch: enabled
    Partitioner: org.apache.cassandra.dht.Murmur3Partitioner
    Schema versions:
        7894d433-6f16-3e83-8d84-ecdb140c7f96: [10.12.72.141, 10.12.72.148, 10.12.74.152, 10.12.74.241]

Stats for all nodes:
    Live: 4
    Joining: 0
    Moving: 0
    Leaving: 0
    Unreachable: 0

Data Centers:
    ppsCI #Nodes: 4 #Down: 0

Database versions:
    4.0.14: [10.12.72.141:7000, 10.12.72.148:7000, 10.12.74.152:7000, 10.12.74.241:7000]

Keyspaces:
    system_schema -> Replication class: LocalStrategy {}
    system -> Replication class: LocalStrategy {}
    system_auth -> Replication class: NetworkTopologyStrategy {ppsCI=3}
    system_distributed -> Replication class: NetworkTopologyStrategy {ppsCI=3}
    system_traces -> Replication class: NetworkTopologyStrategy {ppsCI=3}
    pps -> Replication class: NetworkTopologyStrategy {ppsCI=3}
```

```
root@poojatestcassandra1-az-prod-ci:~# cqlsh -u cassandra -p **** `hostname -i`
Connected to PPS Cassandra restore at 10.12.72.148:9042
[cqlsh 6.0.0 | Cassandra 4.0.14 | CQL spec 3.4.5 | Native protocol v5]
Use HELP for help.
cassandra@cqlsh> LIST USERS ;

 name        | super | datacenters
-------------+-------+-------------
   cassandra |  True |         ALL
 ppsreadonly | False |         ALL
     ppsuser |  True |         ALL

(3 rows)
cassandra@cqlsh> desc KEYSPACES ;

pps     system_auth         system_schema  system_views
system  system_distributed  system_traces  system_virtual_schema
```