Performing Backups

Once the configuration files are setup with the correct prefix and bucket-name, backups through medusa can be safely perfomed.

Cluster backup without PSSH utility

To take a cluster wide backup without pssh utility, it is important to schedule a backup on each and every node of the cluster at the exact same time, only then it will a cluster-wide backup.

Step 1 : Run the backup using the below command.

```
medusa backup --backup-name=<backup name>

Example : medusa backup --backup-name=backup-$(date +%d-%b-%Y-%H-%M) -- This will create backup with name :
backup-17-Apr-2025-01-00
```

Step 2 : Tail the logs and observe for any errors. You would see the details of the logs which will let you know about the steps happening during the whole process like - Snapshot creation, comparing the files and skipping which is already backed-up, uploading the new files to the blob. It will also print out the start and end of the backup.

```
[2025-04-17 01:00:02,273] INFO: Registered backup id backup-17-Apr-2025-01-00
[2025-04-17 01:00:02,273] INFO: Monitoring provider is noop
[2025-04-17 01:00:02,534] INFO: Starting backup using Stagger: None Mode: differential Name: backup-17-
Apr-2025-01-00.
[2025-04-17 01:00:02,535] INFO: Saving tokenmap and schema
[2025-04-17 01:00:03,402] INFO: Resolving ip address 10.12.77.171
[2025-04-17 01:00:03,402] INFO: ip address to resolve 10.12.77.171
[2025-04-17 01:00:03,639] INFO: Saving server version
[2025-04-17 01:00:04,602] INFO: Creating snapshot
[2025-04-17 01:00:06,882] INFO: Listing already backed up files for node 10.12.77.171
[2025-04-17 01:00:07,466] INFO: Backing up myntra_chronicleerp.logtimelookup-
896adb10b84011e9bbe8ef2322b55238
[2025-04-17 01:00:14,886] INFO: Skipping upload of 88 files in myntra_chronicleerp.logtimelookup-
896adb10b84011e9bbe8ef2322b55238 because they are already in storage
[2025-04-17 01:00:14,887] INFO: Backing up myntra_chronicleerp.log-eb72ae50b84011e9bbe8ef2322b55238
[2025-04-17 01:00:33,691] INFO: Skipping upload of 160 files in myntra_chronicleerp.log-
eb72ae50b84011e9bbe8ef2322b55238 because they are already in storage
[2025-04-17 01:00:38,584] INFO: Skipping upload of 16 files in system_schema.columns-
24101c25a2ae3af787c1b40ee1aca33f because they are already in storage
[2025-04-17 01:00:38,585] INFO: Backing up system_schema.functions-96489b7980be3e14a70166a0b9159450
[2025-04-17 01:00:38,585] INFO: Backing up system_schema.aggregates-924c55872e3a345bb10c12f37c1ba895
[2025-04-17 01:00:39,487] INFO: Updating backup index
[2025-04-17 01:00:39,949] INFO: Backup done.
```

Step 3 : You would see the details at the end of log which consists of information like number of files that were skipped, the number of files that were uploaded, total amount of data backed up for that node etc. as shown below.

```
[2025-04-17 01:00:39,949] INFO: - Started: 2025-04-17 01:00:02
                - Started extracting data: 2025-04-17 01:00:04
                - Finished: 2025-04-17 01:00:39
[2025-04-17 01:00:39,949] INFO: - Real duration: 0:00:35.347359 (excludes time waiting for other nodes)
[2025-04-17 01:00:39,950] INFO: - 608 files, 1.68 TB
[2025-04-17 01:00:39,950] INFO: - 152 files copied from host (152 new, 0 reuploaded)
[2025-04-17 01:00:39,950] INFO: - 456 kept from previous backup (backup-17-Apr-2025-01-00)
```

Step 4 : **To check list of backups for the node.**

```
medusa list-backups
```

The output of the above will look something like below

```
 medusa list-backups

[2025-04-17 12:42:38,937] INFO: Resolving ip address
[2025-04-17 12:42:38,938] INFO: ip address to resolve 10.12.77.171
backup-15-Apr-2025-01-00 (started: 2025-04-15 01:00:02, finished: 2025-04-15 01:00:39)
backup-16-Apr-2025-01-00 (started: 2025-04-16 01:00:02, finished: 2025-04-16 01:03:02)
backup-17-Apr-2025-01-00 (started: 2025-04-17 01:00:02, finished: 2025-04-17 01:44:29)
```

Step 5 : **To check the status of backup that was performed**.

This command will give information about, how many nodes have the same backup, what is the status for each node whether it was successfully backed up, the start and end time of the backup, the total size of the backup(accumulative), the total number of files backed up.

```
medusa status --backup-name <backup name>
```

```
medusa status --backup-name backup-17-Apr-2025-01-00
[2025-04-17 12:46:33,190] INFO: Resolving ip address
[2025-04-17 12:46:33,191] INFO: ip address to resolve 10.12.77.171
backup-17-Apr-2025-01-00
- Started: 2025-04-17 01:00:02, Finished: 2025-04-17 01:44:29
- 30 nodes completed, 0 nodes incomplete, 0 nodes missing
- 21524 files, 50.87 TB
```

Step 6 : **To verify the backup, if everything was backed and looks good, run the verification command**.

```
medusa verify --backup-name <backup-name>
```

```
medusa verify --backup-name backup-17-Apr-2025-01-00
[2025-04-17 12:51:07,288] INFO: Resolving ip address
[2025-04-17 12:51:07,289] INFO: ip address to resolve 10.12.77.171
Validating backup-17-Apr-2025-01-00 ...
- Completion: OK!
- Manifest validated: OK!!
```