Percona Backup for MongoDB in Sharded cluster

This document provides the detailed steps for setting up Percona Backup for MongoDB (pbm) in a sharded cluster.

Steps:

1. **Install pbm-agent in all the nodes where mongod daemon is running i.e all nodes in each shard and config servers.**
    a. Install percona-release:

```
yum install https://repo.percona.com/yum/percona-release-latest.noarch.rpm
```

    b. Enable the repository for pbm:

```
percona-release enable pbm release
```

    c. Install pbm-agent package:

```
yum install percona-backup-mongodb
```

2. **Create pbmuser on replSet of each shard and configrs.**

Note: If you create a user through mongos, it will replicate only in config servers and for shards you need to create it separately.

    a. First create a custom role:

```
db.getSiblingDB("admin").createRole({ "role": "pbmAnyAction",
      "privileges": [
         { "resource": { "anyResource": true },
           "actions": [ "anyAction" ]
         }
      ],
      "roles": []
   });
```

    b. Create a user:

```
db.getSiblingDB("admin").createUser({user: "pbmuser",
       "pwd": "secretpwd",
       "roles" : [
          { "db" : "admin", "role" : "readWrite", "collection": "" },
          { "db" : "admin", "role" : "backup" },
```

```
        { "db" : "admin", "role" : "clusterMonitor" },
        { "db" : "admin", "role" : "restore" },
        { "db" : "admin", "role" : "pbmAnyAction" }
      ]
  });
```

3. **Setup the MongoDB connection URI for pbm-agent.**

pbm-agent processes should connect to their localhost mongod with a standalone type of connection.

a. Edit the environment file and the path is located in the pbm-agent.service file.

```
[root@ip-172-31-90-192 ~]# systemctl status pbm-agent
● pbm-agent.service - pbm-agent
   Loaded: loaded (/usr/lib/systemd/system/pbm-agent.service; enabled; vendor preset: disabled)
   Active: inactive (dead) since Thu 2021-09-16 20:28:58 UTC; 6s ago
  Process: 61462 ExecStart=/usr/bin/pbm-agent (code=killed, signal=TERM)
 Main PID: 61462 (code=killed, signal=TERM)
```

```
[root@ip-172-31-90-192 backup]# cat /usr/lib/systemd/system/pbm-agent.service
[Unit]
Description=pbm-agent
After=time-sync.target network.target

[Service]
EnvironmentFile=-/etc/sysconfig/pbm-agent
Type=simple
User=pbm
Group=pbm
PermissionsStartOnly=true
ExecStart=/usr/bin/pbm-agent

[Install]
WantedBy=multi-user.target
```

b. Edit the environment file and specify MongoDB connection URI string for the pbm user to the local mongod node. ()

```
PBM_MONGODB_URI="mongodb://pbmuser:secretpwd@localhost:27017"
```

4. **Set the MongoDB connection URI for pbm CLI.**

The pbm CLI will ultimately connect to the replica set with the PBM control collections. For sharded cluster, set the connection URI to config server replSet.

```
export PBM_MONGODB_URI="mongodb://pbmuser:secretpwd@mongocsvr1:27018,mongocsvr2:27
018,mongocsvr3:27018/?replicaSet=configrs"
```

5. **Configure Remote backup storage.**

There are 3 options available for remote storage –

- S3-compatible storage

```
storage:
  type: s3
  s3:
    region: us-west-2
    bucket: pbm-test-bucket
    prefix: data/pbm/backup
    credentials:
      access-key-id: <your-access-key-id-here>
      secret-access-key: <your-secret-key-here>
    serverSideEncryption:
      sseAlgorithm: aws:kms
      kmsKeyID: <your-kms-key-here>
```

- Filesystem type storage

```
storage:
  type: filesystem
  filesystem:
    path: /data/local_backups
```

- Microsoft Azure Blob storage

```
storage:
  type: azure
  azure:
    account: <your-account>
    container: <your-container>
    prefix: pbm
    credentials:
      key: <your-access-key>
```

a. Create a config file with name pbm_config.yaml
b. Specify the storage information for above 3 methods in the yaml file.
Note: For Filesystem type, give the ownership of pbm to the backup directory and create this directory on all nodes of each shard and config servers.

```
chown pbm:pbm <backup_directory>
```

c. Insert the config file using pbm command.

```
pbm config --file pbm_config.yaml
```

NOTE: For a sharded cluster, run this command whilst connecting to config server replica set. Otherwise connect to the non-sharded replica set as normal.

6. **Start the pbm-agent process.**

Start the pbm-agent process on all nodes of each shard and config servers.

```
systemctl start pbm-agent
```

7. **Check pbm-agent log.**

Use “journalctl -u pbm-agent.service” command to view the logs.

```
journalctl -u pbm-agent.service
-- Logs begin at Tue 2019-10-22 09:31:34 JST. --
Jan 22 15:59:14 akira-x1 systemd[1]: Started pbm-agent.
Jan 22 15:59:14 akira-x1 pbm-agent[3579]: pbm agent is listening for the commands
...
...
```