MongoDB DBA's Guide to Actifio GO

Updated September 22, 2022

Actifio GO

Contents

Preface

The information presented in this guide is intended for users who are familiar with basic Actifio processes and procedures as described in ***Getting Started with Actifio Copy Data Management*** and who are qualified to administer MongoDB databases.

The ActifioNOW Customer Portal

During the configuration and initialization of your Actifio appliance your Actifio representative provided you with a user name and password for the ActifioNOW customer portal.

From the customer portal you can obtain detailed reports about your Actifio appliance as well as search the portal's knowledge base for answers to specific questions.

To log into the ActifioNOW customer portal:

1. Go to: **https://now.actifio.com**
2. When prompted, enter the user name and password provided by your Actifio representative.

To contact an Actifio support representative, send email to: support@actifio.com

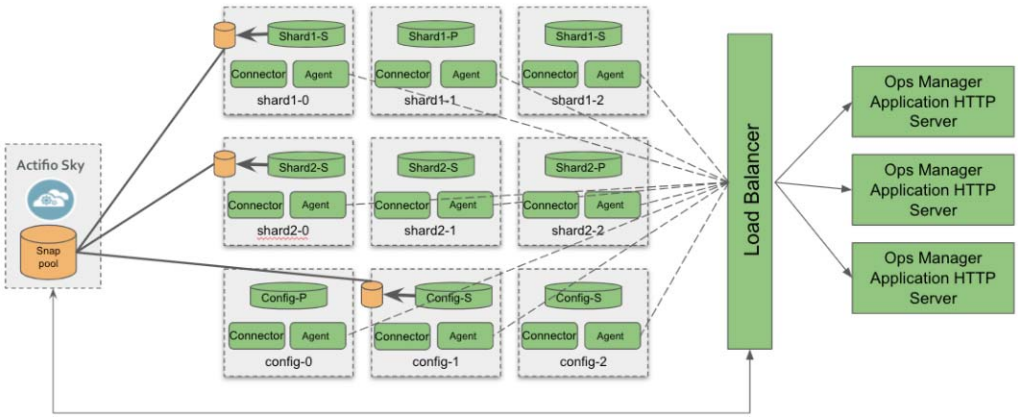
1 Introducing Actifio for MongoDB Sharded and Non-Sharded Clusters

MongoDB is a document-oriented NoSQL database that provides highly scalable, accessible, and efficient queries on large databases. It uses documents and collections for organizing the data, provides scalability through multiple shards, and provides reliability through multiple replica sets.

This section discusses support for protecting and recovering MongoDB sharded and non sharded (with or without arbiter) clusters efficiently where protection comprises one full volume level backup followed by incremental volume level backups and point in time ops log capture while recovery provides point in time restore back to the source or a different target cluster.

MongoDB backup flow

1. The Actifio Connector (lightweight agent) runs on all MongoDB cluster nodes. The appliance communicates with the appropriate nodes depending on the type of request or running job.
2. All communication with Ops Manager happens directly from the Actifio appliance using Ops Manager API calls. Ops Manager communicates with the MongoDB nodes for carrying out operations such as preparing for the application snapshot.
3. MongoDB database backups are initiated by the appliance, invoking Ops Manager APIs to run an application deep discovery.
4. Ops Manager responds with node details, tags, and the volume details of the MongoDB application on those nodes.
5. An individual backup job is started on all the participating nodes of the MongoDB cluster. The appliance communicates with Ops Manager to quiesce the data on disk in preparation for volume snapshots on participating nodes.
6. The application remains quiesced for a brief period while the volume snapshots are captured. Data movement happens from the volume snapshots without impacting the application.
7. Configuration files from all participating nodes are captured and snapshots of the backup are stored together as a consolidated backup of the entire cluster.



Ops Manager Reference Guide:

General installation overview for Ops Manager: `https://www.mongodb.com/docs/ops-manager/current/installation/`

Ops Manager prerequisites:

- Every node of the cluster must have the MongoDB Agent installed.
  `https://www.mongodb.com/docs/ops-manager/current/tutorial/nav/mongodb-agent/`
- API Key pair must be generated with GLOBAL_BACKUP_ADMIN privileges and used when creating the Ops Manager host prior to discovering clusters.:
  `https://www.mongodb.com/docs/ops-manager/current/tutorial/nav/install-application/`

Ops Manager required appsettings:

- `mms.featureFlag.backup.thirdPartyManaged`: must be set to either enabled (recommended) or controlled. If controlled the feature flag must be turned on at the project level in the beta features section for that project.
- `brs.thirdparty.baseOplogFilePath`: must be set with a global directory value such as /oplogs.

Ops Manager oplogs:

- Make sure the oplogs directory defined in `brs.thirdparty.baseOplogFilePath` is created and owned/writable by the MongoDB Agent user on all MongoDB nodes that will be used for backup.
- /oplogs mnt created on all nodes used for backup should be sized to
  (oplog/hr * number of hours to retain in case oplog copying stops) plus a buffer of 20%.
  If the oplog snapshotting process is down for some time, then the tail will have to stop if the disk gets full, creating the risk of losing some oplogs.

Note: *If both csrs and shard data are sharing a node, only one volume for /oplogs is needed and both can use it.*

---

2 Adding the Ops Manager Host and Discovering the MongoDB Cluster

Prerequisites

- MongoDB cluster is set up with Ops Manager.
- All nodes of the cluster are managed by Ops Manager and have the MongoDB Agent installed and can connect to Ops Manager.
- "Backup and monitoring" must be activated under the Ops Manager GUI for all the replica set nodes of the MongoDB cluster.
- All cluster nodes are configured with the backup tag value (key=gcbackupdr). The tag is used to determine which nodes participate in the backup. Assign tag values based on the data center. For example: In us-east1 datacenter the tag key:<value> will be “gcbackupdr”:”us-east1”. This can later be used to ensure backups only run on cluster nodes in the us-east1 datacenter.
- The MongoDB database data volume is LVM managed and the corresponding volume group has at least 20% of free space, on all the nodes.
- Config and shard data that coexist on a single node are stored on separate LVs.
- Actifio AGM and Sky appliance are installed with all required network access to the Ops Manager and MongoDB cluster nodes.
- The Actifio connector is installed on all nodes of the cluster (not needed on Ops Manager). See Install the Actifio Connector on the Host.
- Network connectivity from Sky to the Ops Manager must be open on port 8443 (unless it was changed in the AGM at Manage > Host > Ops Manager Settings).
- Network connectivity for iSCSI from all MongoDB cluster nodes to Sky must be open, as well as connector traffic (port 5106). Reference the firewall rule section in ***Network Administrator’s Guide to Actifio VDP*** for details.

Install the Actifio Connector on the Host

The Actifio Connector, a small-footprint, lightweight service installed on all nodes of the MongoDB cluster, is used to capture an application-consistent copy of MongoDB cluster database. It uses changed block tracking to identify changes to database data for incremental forever capture strategy.

> ***Note:*** *All nodes of the MongoDB Sharded cluster must have both the MongoDB Agent and the Actifio Connector installed. The Actifio Connector is not needed on Ops Manager.*

The Actifio Connector is fully detailed in ***Network Administrator’s Guide to Actifio VDP***.

You can:

- Install the Actifio Connector on a Linux Server via CLI
- Install the Actifio Connector on a Linux Server via AGM

Install the Actifio Connector on a Linux Server via CLI

To install the Actifio Connector on a Linux server:

1. Log on to the Linux server as root.
2. Download the connector using the curl CLI command:
   `curl -O http:///connector-Linux-latestversion.rpm`
3. To check the RPM package before proceeding with installation, run:
   `rpm --checksig connector-Linux-latestversion.rpm`
4. Install the Actifio Connector at /opt/act:
   `sudo rpm -ivh connector-Linux-latestversion.rpm`
5. Verify that the Actifio Connector is running by executing:
   `systemctl status udsagent`

Install the Actifio Connector on a Linux Server via AGM

To install the Actifio Connector on a Linux server using the AGM:

1. Log on to the Linux server as root.
2. Open a browser to the Actifio Resource Center at http://<Sky appliance IP address> and select the Linux Connector icon.
3. Click **OK** in the information dialog and double-click the downloaded file to run the installer.

Defining Actifio Policy Templates and Resource Profiles

The policy template and the resource profile that make up the SLA define the type of data capture to perform and where to store the captured image. Based on the required RPOs and desired backup storage location we will need to create Templates and Profiles. The policy part of the template provides additional backup configuration like enabling logs backups.

A resource profile defines the Sky appliances that are identified as the primary (local) appliance hosting the disk pool to use for snapshots and a remote appliance for remote backup to be used for disaster recovery operations. The profile also defines the OnVault pools to be used to send backup data to an object storage device or cloud offering, such as Google Cloud Storage. You can use an existing template and profile to protect the application or create new template and profile.

The procedures for developing SLAs are detailed in the AGM online help. This chapter provides additional information of value to the MongoDB DBA.

The policy template and the resource profile that make up the SLA define the type of data capture to perform and where to store the captured image. Based on the required RPOs and desired backup storage location we will need to create Templates and Profiles. The policy part of the template provides additional backup configuration like enabling logs backups.

When creating a snapshot policy as part of creating a template you have the option of also capturing its log files at a specified frequency. Details are in Chapter 3, Modifying Protection of a MongoDB Cluster and its Logs.

Discovering and Protecting a MongoDB Cluster Application

Before you can protect a MongoDB cluster, you must add the Ops Manager host and discover the cluster. This requires:

- Add the Ops Manager Host to the AGM
- Discover and Protect the MongoDB Cluster from the App Manager on page 5

Add the Ops Manager Host to the AGM

Add the host to AGM.

1. From the AGM Manage, Hosts page, click **+Add Host**.
2. On the Add Host page:
    - **Name**: Provide the Ops Manager hostname. This must resolve from the appliance.
    - **IP Address**: Provide the Ops Manager IP address and click the + in the right corner.
    - **Appliances**: Select the check box for the Sky Appliance that will manage the data.
    - **Host Type**: Make sure this is MongoDB Ops Manager.
    - **MongoDB Ops Manager Settings**:
        - Update port (optional)
        - Public key (Public keys are from the MongoDB Ops Manager API Key pair.)
        - Private key. (Private keys are from the MongoDB Ops Manager API Key pair.)
3. Click **Add** in the bottom right of the page to add the host.

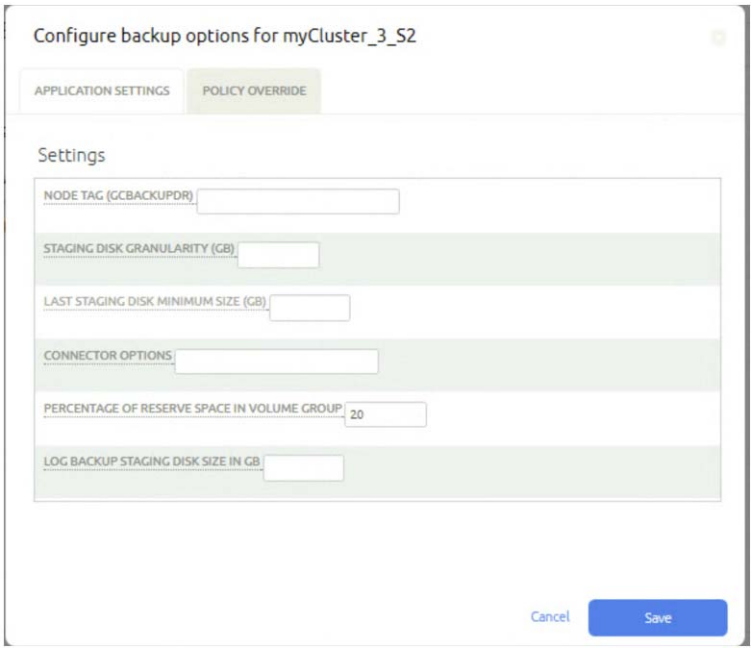
Discover and Protect the MongoDB Cluster from the App Manager

To discover and protect the MongoDB cluster:

1. From the AGM App Manager, Applications page upper right corner, select + **Add Application**.
2. In the Add Application wizard, select **MongoDB**. This opens a five-step onboarding wizard:
    a. In the Discover step, select the Ops Manager host (or click "+ Add Host" to add one if you have not yet) and click **Next** to discover all MongoDB clusters managed by the selected Ops Manager on the prior screen.
    b. In the Select step, select the MongoDB cluster(s) to manage from discovered clusters
    c. In the Manage step, select the cluster to protect and from the drop down list select **Apply SLA**, select a template, select a profile (that you created in Defining Actifio Policy Templates and Resource Profiles) and click **Ok**.
    Click **Next** at the bottom of the page.
    d. In the Configure step, click **Application Settings** (refer to the table below).

Application Setting	Description
NODE TAG (GCBACKUPDR)	Select the appropriate backup tag to determine which node(s) will be used to run backups. ***Note:*** *All cluster nodes should be configured with the backup tag value (key=gcbackupdr). The tag is used to determine which nodes participate in the backup. The same tags are also used on the target cluster during restore for grouping related replicas together for restore/resync action. Assign tag values based on the data center.* ***Note:*** *For multi-shard clusters, there should be at least one replica for each shard with a matching tag (specified here) for the backup job to succeed. If this setting is left empty, one node from each shard will be selected for backup irrespective of the tag value.*

Application Setting	Description
Staging Disk Granularity (GB)	For applications that are under the size of granularity setting that tend to periodically grow this new option is useful to avoid frequent costly FULL backups. Because the staging disk is thin provisioned, there is no initial cost to use a staging disk that is larger than required for immediate use. The values are 0 for No and the Staging Disk Granularity setting for Yes.
Last Staging Disk Minimum Size (GB)	Maximum size of each staging disk when multiple staging disks are used for an application. The default value is 1000GB.
Percentage of Reserve Space in Volume Group	This is needed for volume level backup to determine the required amount of temporary free space in LVM volume group for snapshot. Recommended value is 20%.
Log Backup Staging Disk Size in GB	Optional, recommended to leave blank for dynamic log disk sizing.



3. Click **Save**, then **Next**, then **Finish**.

3 Modifying Protection of a MongoDB Cluster and its Logs

You can change the policy template and resource profile if needed.

Changing the Protection of a MongoDB Cluster

To protect the MongoDB cluster:

1. From the AGM App Manager, Applications list, right-click the database and select **Manage SLA**.
2. On the Manage SLA page, select a template and a resource profile, then click **Apply SLA**.

3. Click **Save**.

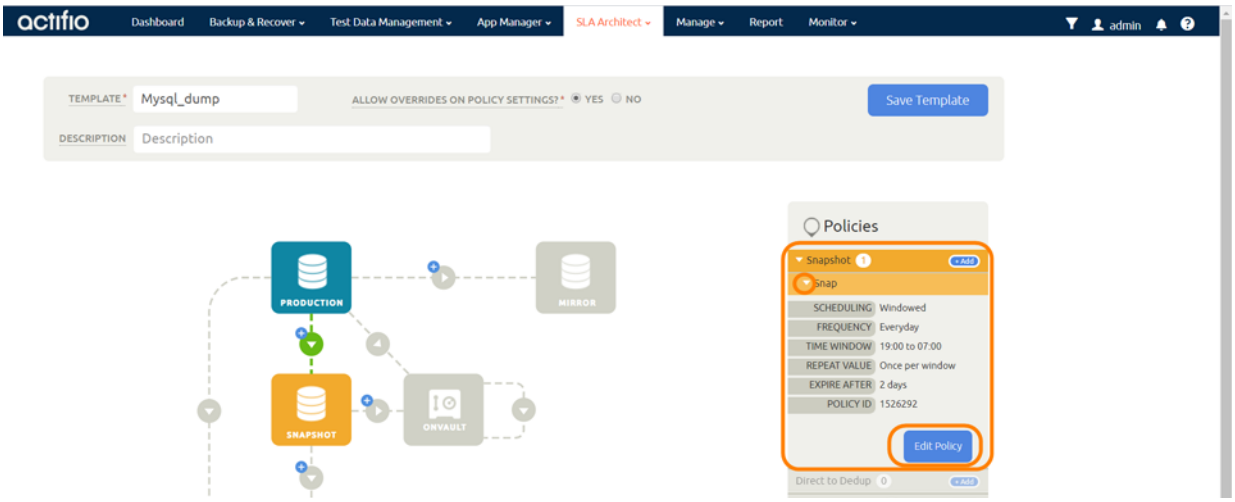
Changing Protection of MongoDB Database Logs

When creating a snapshot policy as part of creating a template you have the option of also capturing its log files at a specified frequency.

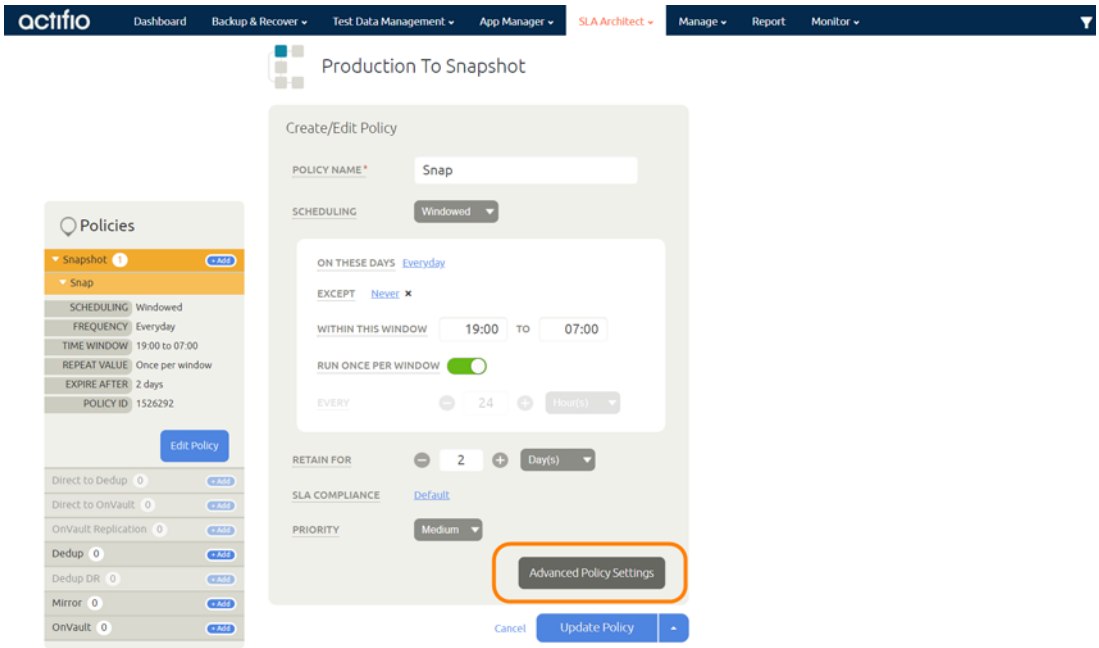
You can replicate database logs to a remote Sky appliance or to an OnVault. You can use the logs at the remote site for any database image within the retention range of the replicated logs. Log replication uses StreamSnap technology to perform the replication between the local and remote Sky appliances; log replication goes directly from the local snapshot pool to the snapshot pool on the remote appliance. For a log replication job to run, there must be a replication policy included in the template, and at least one successful replication of the database must first be completed.

To enable and set up the MongoDB database log backup:

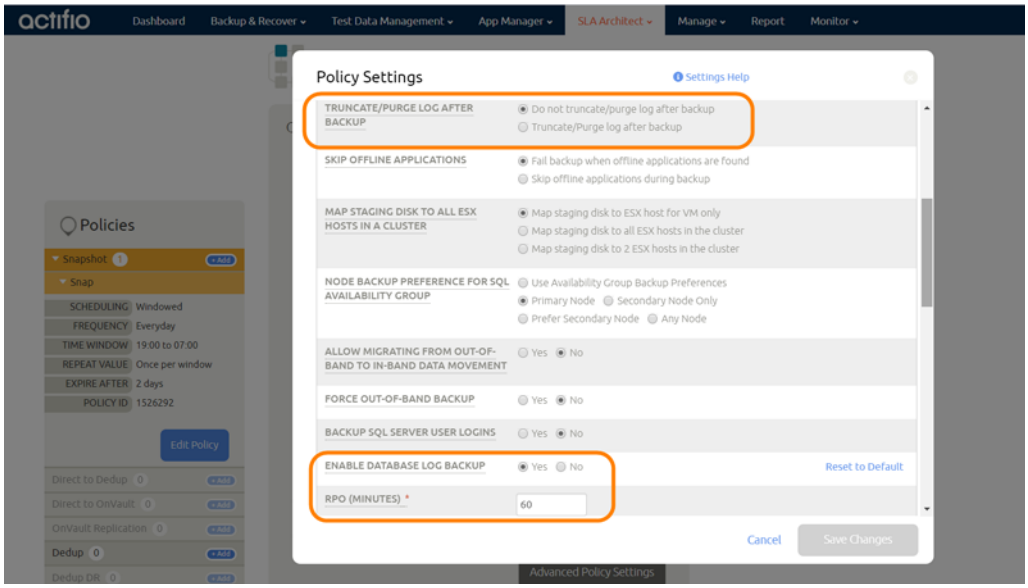
1. From the SLA Architect templates list, right-click the template for MongoDB database protection and click **Edit**.
2. Click the arrow beside the Snapshot policy to open up the details, then click **Edit Policy**.



3. Near the bottom of the Create/Edit Policy page, select **Advanced Policy Settings**.



4. Set the log policy options (you will have to scroll to see them all):
   - Enable **Truncate/Purge log after backup**.
   - Set **Enable Database Log Backup** to **Yes**.
   - For **RPO (Minutes)**, enter the desired frequency of log backup.
   - Set **Log Backup Retention Period (in Days)** for point in time recovery.
   - Set **Replicate Logs (Uses StreamSnap Technology)** to **Yes** if you want to enable StreamSnap replication of log backup to a DR site.
   - Set **Send Logs to OnVault Pool** to **Yes** if you want the database logs to be sent to an OnVault Pool, enabling for point-in-time recoveries from OnVault on another site.



5. Click **Save Changes**.

4 Restoring or Recovering a MongoDB Cluster Database

Restore is supported back to the source cluster or to a new target MongoDB cluster

Prerequisites

- The target cluster has the same topology (number of shards) as the source cluster at the time of the backup.
- All nodes of the cluster are managed by Ops Manager and have the MongoDB Agent installed and can connect to Ops Manager.
- All nodes of the target MongoDB cluster have the Actifio connector installed (the Connector is not needed on Ops Manager).
- The restore will copy data at the volume level by unmounting the volumes on the target cluster nodes. Make sure MongoDB dbpath mountpoint should be exclusively used by MongoDB database processes and no other processes should share them on any replica node.
- LVM logical volumes on the target nodes for restore should be at least as large as the logical volumes on the source at the time of backup.
- Make sure all target cluster nodes are configured with the tag value (key=gcbackupdr). Assign tag values based on the data center.
  For example: In us-east1 datacenter the tag key:<value> will be “gcbackupdr”:”us-east1”

Automated Recovery of a MongoDB Database Cluster to the Source or to a New Target Cluster

Note: *This procedure will perform a physical recovery of the entire data set. The data is copied, overwriting the original data. This can take a long time depending on the amount of data.*

1. From the App Manager, Applications list, right-click the MongoDB database and select **Access**.
2. Select the snapshot to recover, and choose **Restore**.
3. On the restore screen, select the Ops Manager and the target cluster. The default is to overwrite the source cluster.
    - **OPS Manager**: Select the Ops Manager that is managing the target cluster for restore.
    - **Cluster**: Select the cluster for the restore target. Only valid targets with the same topology are presented for selection during a restore operation.
    - **Replace Original application Identity**:
      This is only available when a restore is performed to a new MongoDB cluster on the same appliance where the backup was originally generated.

      Yes replaces the original application reference in application manager and will carry the same application id, jobhistory, backup images, and SLA as the original application.

      No does not replace the original application in application manager. It will be discovered as a new application as a part of the restore job.
    - **Node Actions**: All nodes with the same tag are grouped together in the Node Actions section. A restore or resync action can be assigned to each group. Nodes without any tag are listed under a group with the None tag.
      For each node group, select Restore or Resync under the Actions drop down on the right.

      Restore: A copy of the database and the op log backup is presented to all the nodes selected for restore. Data is explicitly moved to the target datapath and MongoDB runs recovery actively using the target datapath after copy.

      Resync: The nodes selected for resync action do *not* actively participate in restore and recovery, but they are resynced by MongoDB after the recovery.

    Note: *Data movement for recovery consumes a lot of network resources. To avoid slowing down the restore RTO by sending data over a WAN, select only nodes in the same datacenter/region as the appliance for the restore action.*

    Note: *If total number of eligible nodes (i.e. excluding arbiter nodes) that are selected for the restore action is not greater than 50%, then the cluster runs in read only mode post recovery until 50% quorum is reached after resyncing additional nodes.*

4. At the bottom of the page, click **Submit**.