High availability for replica shards

When you enable database replication, Redis Enterprise Software copies your data to a replica node to make your data highly available. If the replica node fails or if the master node fails and the replica is promoted to master, the remaining master node is a single point of failure.

You can configure high availability for replica shards so that the cluster automatically migrates the replica shards to an available node. This process is known as replica high availability or replica_ha (also known as slave_ha or Slave HA in the UI).

An available node:

1. Meets replica migration requirements, such as rack-awareness.
2. Has enough available RAM to store the replica shard.
3. Does not also contain the master shard.

In practice, replica migration creates a new replica shard and copies the data from the master shard to the new replica shard.

For example:

1. Node:2 has a master shard and node:3 has the corresponding replica shard.
2. Either:
   - Node:2 fails and the replica shard on node:3 is promoted to master.
   - Node:3 fails and the master shard is no longer replicated to the replica shard on the failed node.
3. If replica HA is enabled, a new replica shard is created on an available node.
4. The data from the master shard is replicated to the new replica shard.

<table>
<thead>
<tr>
<th>Note:</th>
</tr>
</thead>
</table>
| • Replica HA follows all prerequisites of replica migration, such as rack-awareness.  
• Replica HA migrates as many shards as possible based on available DRAM in the target node. When no DRAM is available, replica HA stops migrating replica shards to that node. |

Configuring high availability for replica shards

Using `rladmin` or the REST API, replica HA is controlled on the database level and on the cluster level. You can enable or disable replica HA for a database or for the entire cluster.

When replica HA is enabled for both the cluster and a database, replica shards for that database are automatically migrated to another node in the event of a master or replica shard failure. If replica HA is disabled at the cluster level, replica HA will not migrate replica shards even if replica HA is enabled for a database.

<table>
<thead>
<tr>
<th>Note:</th>
</tr>
</thead>
<tbody>
<tr>
<td>By default, replica HA is enabled for the cluster and disabled for each database.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Note:</th>
</tr>
</thead>
<tbody>
<tr>
<td>For Active-Active databases, replica HA is enabled for the database by default to make sure that replica shards are available for Active-Active replication.</td>
</tr>
</tbody>
</table>

To enable replica HA for a cluster using `rladmin`, run:
To disable replica HA for a specific database using rladmin, run:

```
rladmin tune db <bdb_uid> slave_ha disabled
```

**Configuration options**

You can see the current configuration options for replica HA with:

```
rladmin info cluster
```

**Grace period**

By default, replica HA has a 10-minute grace period after node failure and before new replica shards are created.

To configure this grace period from rladmin, run:

```
rladmin tune cluster slave_ha_grace_period <time_in_seconds>
```

**Shard priority**

Replica shard migration is based on priority. When memory resources are limited, the most important replica shards are migrated first:

1. `slave_ha_priority` - Replica shards with higher integer values are migrated before shards with lower values.
   
   To assign priority to a database, run:
   
   ```
   rladmin tune db <bdb_uid> slave_ha_priority <positive integer>
   ```

2. Active-Active databases - Active-Active database synchronization uses replica shards to synchronize between the replicas.

3. Database size - It is easier and more efficient to move replica shards of smaller databases.

4. Database UID - The replica shards of databases with a higher UID are moved first.

**Cooldown periods**

Both the cluster and the database have cooldown periods. After node failure, the cluster cooldown period prevents another replica migration due to another node failure for any database in the cluster until the cooldown period ends (default: one hour).

After a database is migrated with replica HA, it cannot go through another migration due to another node failure until the cooldown period for the database ends (Default: two hours).

To configure this grace period from rladmin, run:

- For the cluster:

  ```
  rladmin tune cluster slave_ha_cooldown_period <time_in_seconds>
  ```

- For all databases in the cluster:

  ```
  rladmin tune cluster slave_ha_bdb_cooldown_period <time_in_seconds>
  ```
Alerts
The following alerts are sent during replica HA activation:

- Shard migration begins after the grace period.
- Shard migration fails because there is no available node (sent hourly).
- Shard migration is delayed because of the cooldown period.

Updated: March 11, 2022