

## Feature spec

### Percona XtraDB Cluster Operator

### Asynchronous replication from another site for Disaster Recovery

|             |                                                                                                                       |                       |                               |
|-------------|-----------------------------------------------------------------------------------------------------------------------|-----------------------|-------------------------------|
| JIRA        | <a href="#">K8SPXC-308</a>                                                                                            | Author                | <a href="#">Sergey Pronin</a> |
| Input Data  | <a href="#">Marco Tusa blog post on async replication</a><br><a href="#">Marco Tusa testing results of PXC 8.0.22</a> | GitHub Public Roadmap | <a href="#">Link</a>          |
| Review till | 11/Jun/2021                                                                                                           | Target release        | 1.9.0                         |
| Spec status | Approved                                                                                                              | Spec ticket ID        | <a href="#">PSPEC-19</a>      |

#### Executive Summary

*What is the business value that the feature will deliver*

Disaster recovery protocols are the key to business continuity. Running a MySQL database in a single data center or region might lead to complete data loss. Recovering from backups in another region is an option, but is not enough for low recovery time objectives (RTO).

To solve these problems a new feature will be added to [Percona XtraDB Cluster Operator](#) - asynchronous replication between two clusters. Setting up asynchronous replication from another site (Replica) and exposing PXC nodes (Main).

#### High level functional overview

*Provide high level details about the solution, assumptions and other details*

1. We want to invest into the latest and greatest technologies and will build this feature for MySQL version 8.0 only based on [Automatic Asynchronous Replication Connection Failover](#)
  - a. It is possible to deliver this feature for 5.7 as well (see [here](#)), but it uses a different approach that would increase maintenance costs on our end.
2. This feature handles a manual failover process. This means that we do not introduce new concepts for external orchestrators and monitoring systems.
3. This feature automates the configuration of Source and Replica Percona XtraDB Clusters in Kubernetes. But we keep in mind that either Source or Replica can run outside of Kubernetes and be out of Operators' control. In such a case the feature will still work.
4. We will design this feature with the idea that we will reuse it for our future Percona Server for MySQL Operator based on Percona Server for MySQL.

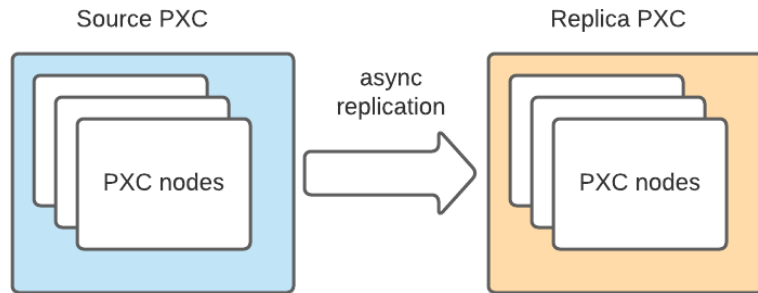
#### Dependencies

*List the milestones on which this specification depends on.*

Release of Percona XtraDB Cluster version 8.0.22 ([RM-882](#), March 18th) to support Automatic Async Repl Failover.  
Also Marco Tusa suggested that 8.0.23 has some enhancements and PXC 8.0.23 should be released in early June.

## Requirements

For requirements and further discussion we assume the following scheme:



| User stories                                                                                   |                                                                                                                                                                                  |          |                            |
|------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------|
| User stories and notes to them                                                                 |                                                                                                                                                                                  |          |                            |
| User story                                                                                     | Notes                                                                                                                                                                            | Priority | JIRA                       |
| Generic stories                                                                                |                                                                                                                                                                                  |          |                            |
| As a user, I want to configure PXC through CR to expose every node with a service              | This is to configure the Source PXC                                                                                                                                              | High     | <a href="#">K8SPXC-308</a> |
| As a user, I want to configure Replica PXC through CR to specify sources for async replication | This is to configure Replica PXC with <code>asynchronous_connection_failover_add_source()</code> and <code>asynchronous_connection_failover_add_source_managed()</code> function | High     | TBD                        |
| As a user, I expect that my Replica cluster is read-only                                       | We cannot control it for the cases when the Replica Cluster is not controlled by the Operator. See <a href="#">RQ 1</a> for QA team.                                             | High     | TBD                        |
| Replication user stories                                                                       |                                                                                                                                                                                  |          |                            |
| As a user, I want to configure                                                                 | User should specify this user on both ends - Source and Replica.                                                                                                                 | High     | TBD                        |

|                                                                                                            |                                                                                                                                          |      |     |
|------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|------|-----|
| replication user through Secret object                                                                     | On Source we also create the grants.                                                                                                     |      |     |
| As a user, I expect that necessary grants for replication user will be created automatically on Source PXC |                                                                                                                                          | High | TBD |
| As a user, I expect that if I change the user in Secret object in Kubernetes grants are changed as well    |                                                                                                                                          | High | TBD |
| <b>Failover stories</b>                                                                                    |                                                                                                                                          |      |     |
| As a user, I expect that promoting Replica to Source is manual                                             |                                                                                                                                          |      | TBD |
| As a user, I expect that promoting Replica to Source disables replication                                  | mysql.replication_asynchronous_connection_failover and mysql.replication_asynchronous_connection_failover_managed tables should be empty | High | TBD |
| As a user, I expect that promoting Replica to Source disables read-only mode                               |                                                                                                                                          | High | TBD |
| As a user, I expect that promoting Replica to Source grants Replication user necessary access              |                                                                                                                                          |      | TBD |
| <b>Documentation stories</b>                                                                               |                                                                                                                                          |      |     |

|                                                                                             |                                                                                                                                                                                                                                                                                                                                                             |      |     |
|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-----|
| As a user, I expect that documentation is updated to describe the replication configuration | <ul style="list-style-type: none"> <li>Describe new user in Secret</li> <li>Describe how to perform a failover</li> <li>Minimal network requirements should be covered (as we need to expose the nodes)</li> </ul>                                                                                                                                          | High | TBD |
| As a product owner, I want to have a blog post written about replication configuration      | Few ideas for blog posts: <ol style="list-style-type: none"> <li>Sausage blog post with how we made the decisions and why</li> <li>Overall review of the feature</li> <li>Highlight use cases:               <ol style="list-style-type: none"> <li>Smooth migration from on-prem to k8s with replication</li> <li>Disaster recovery</li> </ol> </li> </ol> | High | TBD |

### Performance, QA, other

*Define SLAs, RTOs, specific to QA and any other requirements*

#### RQ 1 - QA - Test writable Replica PXC

Users might run Replica PXC outside of the Operator. What kind of issues might we face if something is written to the Replica?

#### RQ 2 - Recovery Point Objectives

We need to communicate in our documentation what kind of RPO the user can expect at certain conditions - DB size, QPS, Network performance between the sites, etc. Even if the outcome is highly parameterized, we expect that RPO in ideal conditions and performance is less than 1 minute.

## Important technical decisions

### ITD 1.1 - UX - add replication user to Secret object

|                                              |                                                                                                                                                                                                                                                              |
|----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>The problem</b>                           | How does the user configure the password for the replication users?                                                                                                                                                                                          |
| <b>Options considered (decision in bold)</b> | <ol style="list-style-type: none"> <li>User to create the user manually</li> <li>Create a separate Kubernetes Secret</li> <li><b>Add one more system user into Users Secret</b></li> <li>Configure user automatically once replication is enabled</li> </ol> |
| <b>Reasoning</b>                             | Operator is all about automation, so we are going to defer option #1. Creating a separate Kubernetes secret for each replication channel might be an option, but introduces additional complexity of user management.                                        |

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     | <p>The decision to create one System user through an already existing mechanism looks like the simplest solution, where all corner cases are already taken care of. The desire to have separate users per replication channel will not be addressed here.</p> <p>Implicit automation (option #4) is something that we are trying to avoid in our Operators.</p>                                                                                                                                                                                                                                                   |
| <b>Details</b>      | <p>New system user “replication” will appear in default my-cluster-secrets. We will allow users to change only the password.<br/>The password and grants will be set automatically.<br/>If the user already exists, we are going to set the replication grants to the user.</p> <pre> apiVersion: v1 kind: Secret metadata:   name: my-cluster-secrets type: Opaque stringData:   root: root_password   xtrabackup: backup_password   monitor: monitory   clustercheck: clustercheckpassword   proxyadmin: admin_password   pmmserver: admin   operator: operatoradmin   <b>replication: repl_password</b> </pre> |
| <b>Consequences</b> | None                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

| ITD 1.2 - UX - add new section to configure replication to cr.yaml |                                                                                                                                                                                                                           |
|--------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| The problem                                                        | How to configure replication on Source and Replica for the cluster controlled by the Operator?                                                                                                                            |
| Options considered (decision in bold)                              | <div>1. Create new Custom Resource in Kubernetes</div> <div>2. <b>Configure using cluster CR (cr.yaml)</b></div>                                                                                                          |
| Reasoning                                                          | Adding a new Custom Resource is a complex process and is also useful if we want to keep the separate “state” somewhere. Replication is just another configuration parameter which can fit existing PXC Customer Resource. |
| Details                                                            | We add the new section <code>spec.pxc.replicationChannels</code>                                                                                                                                                          |
|                                                                    | <code>spec.pxc.replicationChannels</code>                                                                                                                                                                                 |
|                                                                    | Type: array                                                                                                                                                                                                               |

|                                                                                                                                                   |                                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                   | <p>Description: defines the replication configuration for Source or Replica. Array consists of channels.<br/>Default: None (commented out in cr)</p>                                                                    |
|                                                                                                                                                   | <pre>spec.pxc.replicationChannels[].name</pre>                                                                                                                                                                          |
|                                                                                                                                                   | <p>Type: string<br/>Description: defines the name of the channel. Required.<br/>Default: None</p>                                                                                                                       |
|                                                                                                                                                   | <pre>spec.pxc.replicationChannels[].isSource</pre>                                                                                                                                                                      |
|                                                                                                                                                   | <p>Type: boolean<br/>Description: defines if this PXC cluster is Source or Replica<br/>Default: None</p>                                                                                                                |
|                                                                                                                                                   | <pre>spec.pxc.replicationChannels[].sourcesList</pre>                                                                                                                                                                   |
|                                                                                                                                                   | <p>Type: array<br/>Description: defines the list of sources from which Replica should get the data. Ignored if <code>spec.pxc.replicationChannels[].isSource</code> is True, required - if False.<br/>Default: None</p> |
|                                                                                                                                                   | <pre>spec.pxc.replicationChannels[].sourcesList[].host</pre>                                                                                                                                                            |
|                                                                                                                                                   | <p>Type: string<br/>Description: host name or IP-address of the source. Required.<br/>Default: None</p>                                                                                                                 |
|                                                                                                                                                   | <pre>spec.pxc.replicationChannels[].sourcesList[].port</pre>                                                                                                                                                            |
|                                                                                                                                                   | <p>Type: int<br/>Description: port of the source<br/>Default: 3306</p>                                                                                                                                                  |
|                                                                                                                                                   | <pre>spec.pxc.replicationChannels[].sourcesList[].weight</pre>                                                                                                                                                          |
|                                                                                                                                                   | <p>Type: int<br/>Description: weight of the source<br/>Default: 100</p>                                                                                                                                                 |
| <p>Example Source:</p> <pre>spec:   pxc:     replicationChannels:       - name: pxc1_to_pxc2         isSource: true</pre> <p>Example Replica:</p> |                                                                                                                                                                                                                         |

|                     |                                                                                                                                                                                                                                                                                                       |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     | <pre>spec:   pxc:     replicationChannels:       - name: pxc1_to_pxc2         isSource: false         sourcesList:           - host: pxc1.source.percona.com             port: 3306             weight: 100           - host: pxc2.source.percona.com           - host: pxc3.source.percona.com</pre> |
| <b>Consequences</b> | None                                                                                                                                                                                                                                                                                                  |

| ITD 1.3 - UX - Expose every PXC node as a service |                                                                                                                                                                                  |
|---------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| The problem                                       | For Replica cluster to connect to Source every PXC node in Source cluster should be exposed.                                                                                     |
| Options considered (decision in bold)             | <div>1. User to create Service objects manually</div> <div>2. <b>Configure exposure through CR</b></div>                                                                         |
| Reasoning                                         | We run the Operator which should simplify the operations, so we want to do it with CR.                                                                                           |
| Details                                           | We are going to add a new section under spec.pxc - expose.                                                                                                                       |
|                                                   | <div>spec.pxc.expose.enabled</div>                                                                                                                                               |
|                                                   | <div>Type: boolean</div> <div>Description: enables or disables the exposure of the PXC nodes</div> <div>Default: false</div>                                                     |
|                                                   | <div>spec.pxc.expose.type</div>                                                                                                                                                  |
|                                                   | <div>Type: inherited from Kubernetes Service spec.type</div> <div>Description: <a href="#">Service type</a> that will be used</div> <div>Default: ClusterIP</div>                |
|                                                   | <div>spec.pxc.expose.loadBalancerSourceRanges</div>                                                                                                                              |
|                                                   | <div>Type: array</div> <div>Description: network prefix ranges whitelisted for the loadbalancer. Requires spec.pxc.expose.type to be LoadBalancer</div> <div>Default: None</div> |
| <div>spec.pxc.expose.annotations</div>            |                                                                                                                                                                                  |

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     | <div data-bbox="414 210 1419 331"> <p>Type: section<br/> Description: annotations applied to the Service object<br/> Default: None</p> </div> <p>Example:</p> <pre>spec:   pxc:     expose:       enabled: true       type: LoadBalancer       loadBalancerSourceRanges:         - 10.0.0.0/8     annotations:       networking.gke.io/load-balancer-type: "Internal"</pre> <p>This will create the internal LoadBalancer per each PXC node.</p> |
| <b>Consequences</b> | <p>We will reuse the same expose section structure for other components in the future. For example, we will replace serviceAnnotations with expose.annotations. This will be handled by another spec: <a href="#">PSPEC-20</a> and define the deprecation policy.</p>                                                                                                                                                                            |