

# Failover Settings (Zookeeper)

## On this page

- [Failover Basics](#)
- [How Failover Works](#)
- [Create an Aspire Failover Installation](#)
  - [Install a ZooKeeper server](#)
  - [Install Aspire in each server](#)
  - [Failover configuration in each Aspire server](#)
  - [Start Aspire servers in failover mode](#)
  - [Verify your installation](#)
- [How Scheduled Crawls Work in a Failover Environment](#)
- [Security Concerns](#)

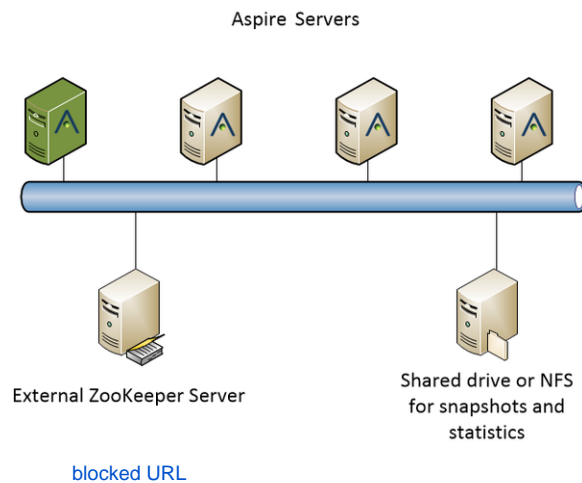
## Failover Basics

Aspire introduces failover features over multiple Aspire servers for content source crawls. For this, Aspire uses Apache ZooKeeper to synchronize configurations (content sources and workflow applications) and coordinate and resume failed crawls, among several Aspire servers.

The failover feature in Aspire intends to maximize the stability of the Content Sources crawls in Distributed environment. When an Aspire server is running a crawl for a content source and for any reason it crashes, all other Aspire Servers connected to the same ZooKeeper and running the same crawl, will notice the fail and will restore and pending crawl items from it to be re-processed.

## How Failover Works

### Step 1. A crawl is running in one Aspire server.



**Step 2. The server running the content source failed so the crawl items are restored in the other servers.**



Aspire synchronizes the content sources configuration and the workflow libraries, so if you create a content source or a library in one server the other one will install that same content source or library.

## Create an Aspire Failover Installation

### Install a ZooKeeper server

By default Aspire starts an embedded ZooKeeper server to run in stand-alone mode. You can use this same server but the failover will only work if that Aspire server is always up and running.

We recommend using external ZooKeeper servers (cluster) so that the uptime of the overall failover functionality can be maximized with ZooKeeper failover too.

For ZooKeeper installation instructions, go to: <http://zookeeper.apache.org/doc/trunk/zookeeperStarted.html>

Also check the Zookeeper machine requirements at: <http://zookeeper.apache.org/doc/r3.1.2/zookeeperAdmin.html#Single+Machine+Requirements>

## Install Aspire in each server

Follow the steps at [Download and Install](#) in each Server to install Aspire

## Failover configuration in each Aspire server

Once you have installed Aspire in each server, edit the **config/settings.xml** file in each distribution folder.

For each Aspire Server make the following change to the `<configAdministration>` section of the **settings.xml** file

```
<zookeeper enabled="false" libraryFolder="config/workflow-libraries" root="/aspire" updatesEnabled="false">
```

to

```
<zookeeper enabled="true" libraryFolder="config/workflow-libraries" root="/aspire" updatesEnabled="true">
```

Uncomment the line with:

```
<!-- <externalServer>127.0.0.1:2182,127.0.0.1:2183,127.0.0.1:2181</externalServer> -->
```

And write the zookeeper server that you have installed as follows:

```
<externalServer> host:port </externalServer>
```

If you are using a cluster of zookeeper servers separate each server with a comma:

```
<externalServer> host1:port1, host2:port2, host3:port3, ... </externalServer>
```

By default if no external server is specified, Aspire will start an embedded ZooKeeper server on the port specified in the **<clientPort>** tag, and it will not be connected to any other ZooKeeper. This is the default for non-failover installations.

## Start Aspire servers in failover mode

1. When you are going to start the Aspire servers, make sure the first server you start is the one with the correct configuration (content sources and workflow libraries) because any previous data stored in ZooKeeper will be replaced with this server's configuration. All subsequent Aspire servers started after the first will replace their own configurations (content sources and workflow libraries) with the one stored in ZooKeeper (set by the first Aspire Server loaded).
2. To avoid unwanted loss of configuration, make sure the first server you start has the content sources and libraries that all of the other servers should share.
3. To start each Aspire server, execute: **bin\aspire.bat** or **bin\aspire.sh** (for Linux servers).

## Verify your installation

1. After you have started all of the Aspire servers, open a browser and go the Home UI of any Aspire server. By default, the UI address can be accessed by browsing to: <http://aspire-server-1:50505>
2. From the Home UI of Aspire, create, configure, and save any content source.
3. Wait until the content source is successfully loaded.

4. Open the Home UI of the rest of the servers and make sure they all have the same content source configured.

## How Scheduled Crawls Work in a Failover Environment

---

If you configured a content source to crawl using a schedule of time, the same configuration will be applied to all servers. And when the time comes to crawl by the schedule, **only one server** will perform the crawl start trigger, the others will eventually start the Distributed Crawl when all the initial setup is done.

## Security Concerns

---

The Failover feature does not enforce Aspire to use security among all Aspire servers, it is completely possible to have a scenario where you have three Aspire Servers in a cluster with only two of them have security access restrictions, the third one will not know about this and will perform its failover crawls without caring about it, so you would end up with an unsecured server crawling your sensitive data.

It is recommended to configure all Aspire servers with security if your project requirements demand it. Go to [Aspire Security](#) for more details about how to configure security access restrictions to Aspire.