

Clustering Artifactory

Active/Active Architecture

Artifactory HA is an Active/Active clustered installation of Artifactory that provides a full set of true High Availability features and is supported with an Artifactory [Enterprise License](#).

For full details please refer to [Artifactory High Availability](#).

Active/Passive Architecture

Overview

Artifactory clustered Active/Passive architecture provides fast disaster recovery and can be implemented in one of the following two ways:

- [Deployment on fault-tolerant storage](#) (strongly recommended)
- [Periodic cross-server data sync](#).

Deployment on Fault-tolerant Storage

Using a fault-tolerant disk mounted on another machine allows for a very short MTR (Mean Time to Recovery) in case the "active" server goes down. If Artifactory is deployed on a NAS or SAN the "passive" machine can immediately mount the storage, bootstrap Artifactory from it and start accepting requests in place of the originally "active" machine that is has gone down.

Page Contents

- [Active/Active Architecture](#)
- [Active/Passive Architecture](#)
 - [Overview](#)
 - [Deployment on Fault-tolerant Storage](#)
 - [Cross-server Data Synchronization](#)
 - [Synchronizing the Data and Configuration Directories](#)
 - [Synchronizing the Database](#)
 - [Time Synchronization on the Standby Server](#)

To set this up quickly and efficiently, we recommend using the built-in Virtual Machine Failover feature offered by virtualization software providers as follows:

1. Create a VM image that runs the Artifactory startup script and mounts the auxiliary storage.
2. The storage should contain the full Artifactory installation along with the data in a location defined as \$ARTIFACTORY_HOME.
3. Use the VM image on two Virtual Machines and have Artifactory running on one machine while the other machine is readily available as a failover target by the virtualization monitor.

Cross-server Data Synchronization

If deployment on fault-tolerant storage, as described in the previous section, is not possible (or if redundancy is required), fault-tolerance can be achieved by correctly replicating the data folder to a warm standby server.

The setup of an up-to-date passive replication server for the active Artifactory server requires database replication and synchronization of file system directories.

Synchronizing the Data and Configuration Directories

To synchronize the data and configuration directories you need to run `rsync` on `$ARTIFACTORY_HOME/data` and `$ARTIFACTORY_HOME/etc`.

This can be done by running the `rsync` command on `$ARTIFACTORY_HOME` while excluding the directories that are not required as follows:

```
rsync -vvah --del --progress --log-file=/home/replication/replication.log --exclude-from=rsync-excludes.txt \
artifactory@active-artifactory-host:$ARTIFACTORY_HOME/ $ARTIFACTORY_HOME/
```

For the above example the `rsync-excludes.txt` file appears as follows:

```
/work/
/data/tmp*/
/data/cache/
/logs/
```



rsync

The `rsync` should be executed from the passive stand-by server

Synchronizing the Database



Database Replication

Database replication must run **before** executing `rsync`.

The procedure to synchronize a database varies between the different database vendors. Please refer to the relevant documentation for your specific database.

For example, instructions on how to synchronize with MySQL can be found in the MySQL documentation for [How to Set Up Replication](#).

It is also possible to use a full dump/restore procedure on the database to synchronize the database and filestore state. In this case, we recommend that you perform the dump in a single routine along with `rsync` (in case of File System Storage Types).

Time Synchronization on the Standby Server

It is very important that the metadata stored in the database and the data stored on the file system are synchronized on the standby server.

A straightforward way to achieve this, is to make sure that the database synchronized is in a state that is **prior** to the file system (data/filestore) state.

This allows you to:

- Make a database dump before executing the file system sync,
- Activate database replication on demand just before executing `rsync`.

Since the sync operations are not atomic, there may be a gap between the data from `rsync` and data from database replication.



1. The snapshot time that Artifactory is set to is the database replication time.
2. Items synced to the file system which have no representation in the database can be purged by clicking on **Prune Unreferenced Data** in the **Admin** tab and then **Advanced | Maintenance** in the Artifactory configuration.