

FTP Server Setup for Fast Backup and Restore

This is the best practices guide for setting up an FTP Server for use with ClustrixDB Fast Backup and Restore, including how to avoid bottlenecks and optimizing for maximum throughput.

ClustrixDB implements fast backup and restore as a binary backup mechanism which works at the row level. Each ClustrixDB node sends its data directly to the FTP backup target in parallel, eliminating bottlenecks and allowing backup speed to scale with cluster size. Similarly for restore, the initiating node coordinates with other participating nodes in parallel to read from the dump file and restore replicas.

ClustrixDB also supports SFTP with password-base authentication.

For more information see [ClustrixDB Fast Backup and Restore](#).

Choosing an FTP Server

- Clustrix recommends using [vsftpd](#) running on Linux. This FTP software has undergone the most testing by Clustrix QA and is used by the majority of ClustrixDB users.

Configuring the FTP Server

- We recommend using a dedicated FTP server for ClustrixDB backup/restore. If that is not feasible we recommend taking steps to ensure the FTP server isn't being used for other operations during the backup/restore process.
- The FTP server should be used on passive mode as ClustrixDB Fast Backup/Restore does not support active mode.
- Configure the FTP server to allow many concurrent connections from the Clustrix user account otherwise it is possible to bottleneck the backup/restore.
 - if using vsftpd, set MaxClients to be ≥ 3
- Provide sufficient I/O and storage capacity. Fast I/O is essential to handle the concurrency of the backup/restore process.
 - Using a hardware RAID controller with RAID10 is recommended to increase read and write I/O performance.
 - If I/O becomes a bottleneck, the backup/restore will take longer to complete. Since the backup/restore process pins BigC (our garbage collector), if it runs too long it could introduce latencies in the database or cause the disks to become full.
- Use 10Gb or bonded 1Gb ethernet connectivity between the FTP server and the Ethernet switch connected to your ClustrixDB nodes in order to achieve maximum network bandwidth.
- Ensure the backup user is chrooted to their home directory.