



# Database

Let's learn how Chubby uses a database for storage.

## We'll cover the following ^

- Backup
- Mirroring

Initially, Chubby used a replicated version of Berkeley DB ([https://www.usenix.org/legacy/event/usenix99/full\\_papers/olson/olson.pdf](https://www.usenix.org/legacy/event/usenix99/full_papers/olson/olson.pdf)) to store its data. Later, the Chubby team felt that using Berkeley DB exposes Chubby to more risks, so they decided to write a simplified custom database with the following characteristics:

- Simple key/value database using write-ahead logging and snapshotting.
- Atomic operations only and no general transaction support.
- Database log is distributed among replicas using Paxos.

## Backup#

For recovery in case of failure, all database transactions are stored in a transaction log (a write-ahead log). As this transaction log can become very large over time, every few hours, the master of each Chubby cell writes a snapshot of its database to a GFS server in a different building. The use of a separate building ensures both that the backup will survive building damage, and that the backups introduce no cyclic dependencies in the system; a GFS cell in the same building potentially might rely on the

Chubby cell for electing its master.



- Once a snapshot is taken, the previous transaction log is deleted. Therefore, at any time, the complete state of the system is determined by the last snapshot together with the set of transactions from the transaction log.
- Backup databases are used for disaster recovery and to initialize the database of a newly replaced replica without placing a load on other replicas.

## Mirroring#

Mirroring is a technique that allows a system to automatically maintain multiple copies.

- Chubby allows a collection of files to be mirrored from one cell to another.
- Mirroring is commonly used to copy configuration files to various computing clusters distributed around the world.
- Mirroring is fast because the files are small.
- Event mechanism informs immediately if a file is added, deleted, or modified.
- Usually, changes are reflected in dozens of mirrors worldwide under a second.
- Unreachable mirror remains unchanged until connectivity is restored. Updated files are then identified by comparing their checksums.
- A special “global” cell subtree `/ls/global/master` that is mirrored to the subtree `/ls/cell/replica` in every other Chubby cell.
- Global cell is special because its replicas are located in widely separated parts of the world. Global cell is used for:
  - Chubby’s own access control lists (ACLs).
  - Various files in which Chubby cells and other systems advertise

their presence to monitoring services.



- Pointers to allow clients to locate large data sets such as Bigtable cells, and many configuration files for other systems.

← Back

Caching

Next →

Scaling Chubby



Mark as Completed



Report an Issue