TarantoolProposal

Tarantool/Box – efficient in-memory data store. http://tarantool.org

Background

Tarantool/Box, or simply Tarantool, is a high performance key/value storage server. The code is available for free under the terms of BSD license. Supported platforms are GNU/Linux and FreeBSD.

The server maintains all its data in random-access memory, and therefore can handle read requests blazingly fast. At the same time, a copy of the data is kept on non-volatile storage (a disk drive), and inserts and updates are performed atomically.

To ensure atomicity, consistency and crash-safety of the persistent copy, a write-ahead log (WAL) is maintained, and each change is recorded in the WAL before it is considered complete.

If update and delete rate is high, a constantly growing write-ahead log file (or files) can pose a disk space problem, and significantly increase time necessary to restart from disk. A simple solution is employed: the server can be requested to save a concise snapshot of its current data. The underlying operating system's “copy-on-write” feature is employed to take the snapshot in a quick, resource-savvy and non-blocking manner. The “copy-on-write” technique guarantees that snapshotting has minimal impact on server performance.

Tarantool supports replication. Replicas may run locally or on a remote host. Tarantool replication is asynchronous and does not block writes to the master. When or if the master becomes unavailable, the replica can be switched to assume the role of the master.

Tarantool is lock-free. Instead of the underlying operating system's concurrency primitives, Tarantool uses cooperative multitasking environment to simultaneously operate on thousands of connections. While this approach limits server scalability to a single CPU core, in practice it removes competition for the memory bus and sets the scalability limit to the top of memory and network throughput. CPU utilization of a typical highly-loaded Tarantool server is under 10%.

The software is production-ready. Tarantool has been developed and is actively used at Mail.Ru one of the leading Russian web content providers. At Mail.Ru, the sowtware serves the “hottest” data, such as online users and their sessions, online application properties, the map between users and their serving shards.

To conclude, Tarantool/Box is ideal for highly volatile and/or highly accessed Web data. With Tarantool, performance overhead on serving data is minimal: a single server can easily deal with tens or even hundreds of thousands of requests per second. Snapshots can be made when Web user activity is at its lowest, for example at night, and thus add little or no restraint on the top throughput of the system. If the master becomes unavailable, a replica can assume the role of the master with minimal downtime.

Core Developers

Konstantin Osipov - Mail.Ru Yuri Vostrikov - Mail.Ru Alexander Calendarev - ISOEMO

Reliance on salaried developers

Development community

How it fits under Apache Software Foundation umbrella