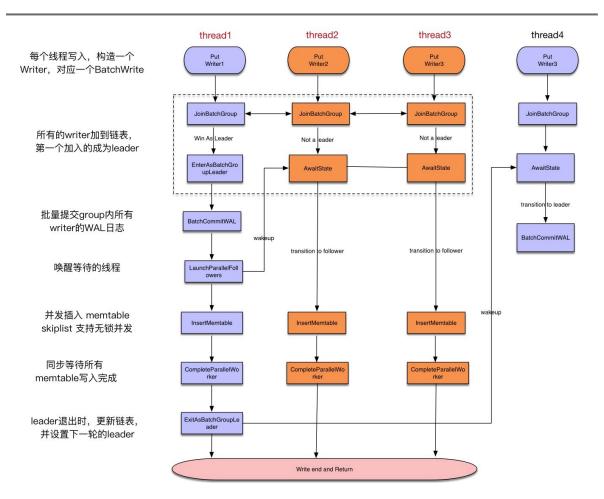
zhuanlan.zhihu.com

RocksDB 写入流程详解

张友东关注他

2 minutes

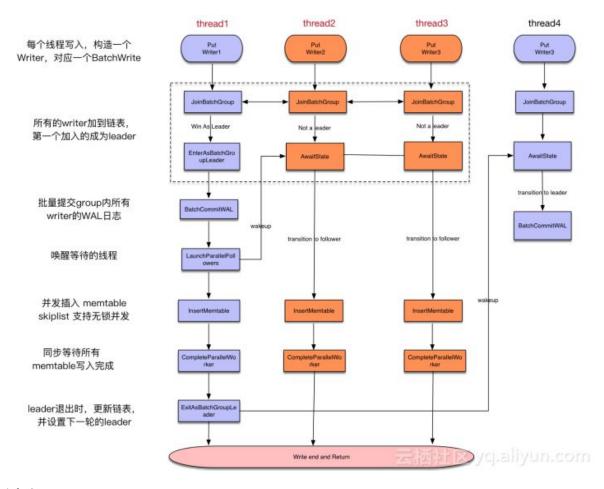


- 最初的写入流程,继承自 leveldb,多个 写线程组成一个 group, leader 负责 group 的 WAL 及 memtable 的提交,提交完后唤醒所有的 follwer,向上层返回。
- 支持 allow_concurrent_memtable_write 选项,在1的基

1 of 3 9/8/2022, 12:30 PM

础上, leader 提交完 WAL 后, group 里所有线程并发写 memtable。原理如下图所示,这个改进在 sync=0的时候,有3倍写入性能提升,在 sync=1时,有2倍性能提升,参考Concurrent inserts and the RocksDB memtable

支持 enable_pipelined_write 选项,在2的基础上,引入流水线,第一个 group 的 WAL 提交后,在执行memtable 写入时,下一个 group 同时开启,已到达Pipeline 写入的效果。



编辑于 2018-01-28 20:11

文章被以下专栏收录

2 of 3 9/8/2022, 12:30 PM



3 of 3