

1. 修改说明

该项目的目的是优化 log-structured merge tree 的写放大，同时在不增加额外设备 (例如 NVM) 的情况下需要兼顾 range query 和 point query 性能。

其思想包括以下三个方面

- 1) 减少 key-value pair 参与内部归并排序的频率
- 2) 尽可能保持每一层 key-value pairs 的有序性，如果保持不了，则保持索引中 keys 的有序性
- 3) 减少小文件的产生，因为这不利于写

该项目对 leveldb 增改了约 2000 行代码，并且进行了初步的实验。选择的实验设备包括 CPU i5-9300H, WD SSD 1TB(560MB/s read, 530MB/s write)。该实验中插入了一百万随机的键值对，key 是一个随机数字，value 的大小为 4096 bytes，该初步实验结果如表 1所示。

表 1: 实验结果

	LevelDB	我修改后的 LevelDB
用户写入数据量	4,151,048,089 bytes	4,151,761,429 bytes
总写入数据量	35,420,564,890 bytes	16,560,785,601 bytes
写放大	8.53	3.99
总花费时间	189.29s	111.36s

很不幸的是，该项目思想已经与已有论文 [1] 发生了冲突，这在我写完了才发现。不过 Yao [1] 并未公布其代码，因此可以拿我写的来印证其思想。

参考文献

[1] Yao, Ting, Jiguang Wan, Ping Huang, Xubin He, Qingxin Gui, Fei Wu, and Changsheng Xie. "A light-weight compaction tree to reduce I/O amplification toward efficient key-value stores." In Proc. 33rd Int. Conf. Massive Storage Syst. Technol.(MSST), pp. 1-13. 2017.