The LSN Concept in InnoDB and RocksDB

Mijin An
meeeeejin@gmail.com

VLDB Lab.

LSN in InnoDB

- Records within the InnoDB redo log are identified via a LSN
- Each page has an LSN (`FIL_PAGE_LSN`)
- This LSN simply shows the **byte offset**:
  - So, when you write 100 bytes to the log buffer, the LSN is increased by 100
- When a **mini-transaction commits**, it writes its own log records to the log buffer, **increasing the LSN**

- The meanings of LSN:
  - **The total amount of redo log writes**
  - **Checkpoint location**
  - **Page version**

```
---
LOG
---
Log sequence number 78939567500
Log flushed up to   78937301748
Pages flushed up to 74300027436
Last checkpoint at  74286584574
0 pending log flushes, 0 pending chkp writes
10089 log i/o's done, 1.94 log i/o's/second
```

LSN in RocksDB

- Each record (KV) in RocksDB has a LSN starting with the first 0, **adding 1 to each write**
- The LSN that can be seen from outside RocksDB is incremented by `WriteBatch`:

```
WriteBatch :=
    sequence: fixed64
    count: fixed32
    data: record[count]
```

  - `WriteBatch`
    - A collection of **atomic operations** of RocksDB **write** operations (`Put(k,v), Merge(k,v), Delete(k), SingleDelete(k)`)
    - Each of these records has a binary string representation:

```
Sequence(0);NumRecords(3);Put(a,1);Merge(a,1);Delete(a);
```

    - `WriteBatch` follows the ACID feature (All or nothing; Isolation from each other)

Reference

[1] Jim Gray and Andreas Reuter, “Transaction Processing: Concepts and Techniques”, Morgan Kaufmann, San Mateo, CA (1993)

[2] “RocksDB v6.14 Document: Two Phase Commit“, https://www.bookstack.cn/read/rocksdb-6.14-en/97444c76bda59850.md

[3] Facebook, “facebook/rocksdb”, GitHub, https://github.com/facebook/rocksdb/blob/master/db/write_batch.cc

[4] “RocksDB transaction realizes TransactionDB analysis”, ProgrammerSought, https://www.programmersought.com/article/51741105962/

[5] “InnoDB Redo Log”, MariaDB, https://mariadb.com/kb/en/innodb-redo-log/

[6] “MySQL (InnoDB analysis): 39---Transaction (implementation of transaction: redo log (redo log))”, ProgrammerSought, https://www.programmersought.com/article/14385450123/