

MariaDB Developer UnConference
April 9-10th 2017
New York



MyRocks in MariaDB

Why and How

Sergei Petrunia
sergey@mariadb.com



What is MyRocks?

What is MyRocks



- RocksDB + MySQL = MyRocks
- LSM architecture
- Better compression
- Better IO efficiency
- Write optimizations
 - Bulk load
 - No-read writes
- Used and verified at Facebook

MyRocks needs adoption and community



- Currently, it is in `github.com/facebook/mysql-5.6`
- No binaries
- No packages
- Very much “in-house” experience
 - Special branch of MySQL
 - Special way to compile
 - Special command to run tests
 - ...

MyRocks + MariaDB = ♥



- New technology



MariaDB

- Adoption
- Packaging
- Community
- MariaDB features



1. Tasks for porting MyRocks
2. Current status
3. Future



- Port MyRocks
from **facebook/mysql-5.6** to **MariaDB**
- Do builds and packages
- Catch a MariaDB release train



- **Port MyRocks**
from [facebook/mysql-5.6](https://facebook.com/mysql-5.6) to MariaDB
- Do builds and packages
- Catch a MariaDB release train

Porting MyRocks



SQL layer

MyRocks

RocksDB

Porting MyRocks



SQL layer

More
counters

Extra
range info

Read-free
replication

Gap lock
use detector

...

MyRocks

counter++

Bloom
filter use

...

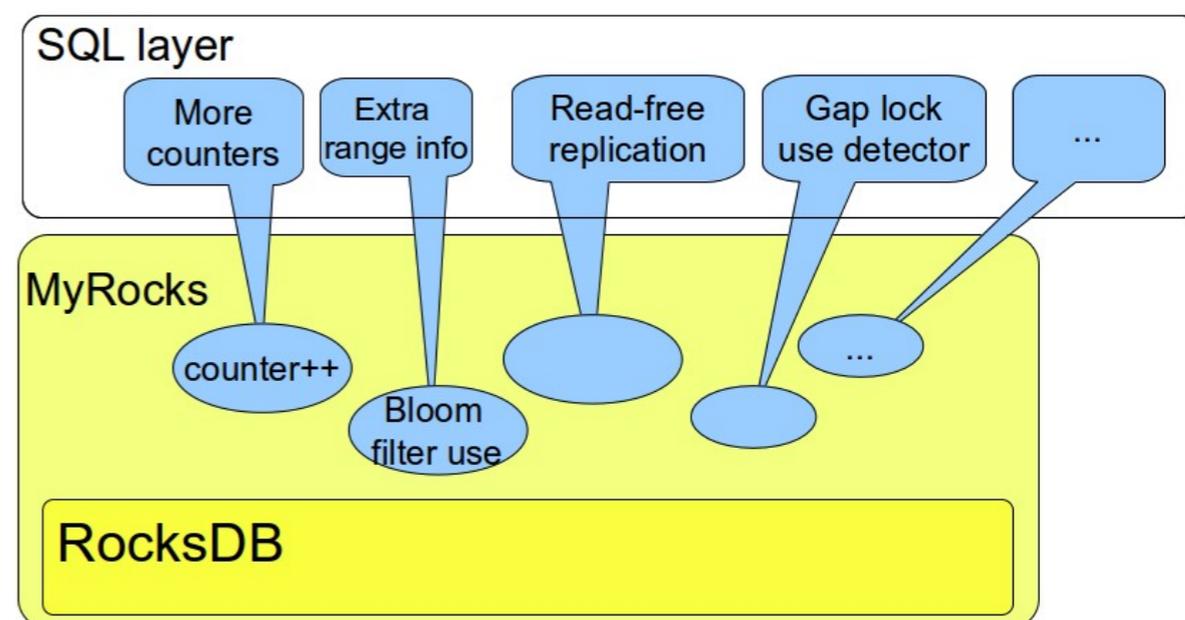
RocksDB

- MyRocks is not just storage/rocksdb/*

Porting MyRocks



- Resolve the SQL layer differences
- Resolve MariaDB vs MySQL differences
 - Index Condition pushdown
 - Extended Keys
 - Group Commit
 - EXPLAIN output changes
 - Extra/missing warnings
 - ...





- Port MyRocks
from **facebook/mysql-5.6** to **MariaDB**
- **Do builds and packages**
- Catch a MariaDB release train

Do builds and packages



- Facebook/mysql-5.6 builds statically
- We want `ha_rocksdb.so`
- We want packages
 - deb, rpm, bintar, windows zip, ...
- RocksDB uses compression libraries
 - Need to link against them
 - Need to have proper dependencies



- Port MyRocks
from **facebook/mysql-5.6** to **MariaDB**
- Do builds and packages
- **Catch a MariaDB release train**



Get into a MariaDB Release before it is declared stable

- MariaDB 10.0:
 - Alpha: 12 Nov 2012
 - Stable: 31 Mar 2014
- MariaDB 10.1 (Stable)
 - Alpha: 30 Jun 2014
 - Stable: 17 Oct 2015
- **MariaDB 10.2 (RC)**
 - **Alpha: 18 Apr 2016**
 - **RC: 17 Feb 2017**
 - **Stable: soon**
- **MariaDB 10.3**
 - **“alpha soon”**



- Plugins declare their maturity
 - unknown, experimental, alpha, beta, gamma, stable
- `mysqld --plugin-maturity=level`
 - will not load plugins less mature than *level*.
- This is how MyRocks can get into MariaDB 10.2
 - Plugin with maturity=alpha.



Current status



- **MariaDB 10.2.5 RC2 includes MyRocks**
- Features
- Packaging
- Quality



- **MariaDB 10.2.5 RC2 includes MyRocks**
- **Features**
- Packaging
- Quality

Feature status

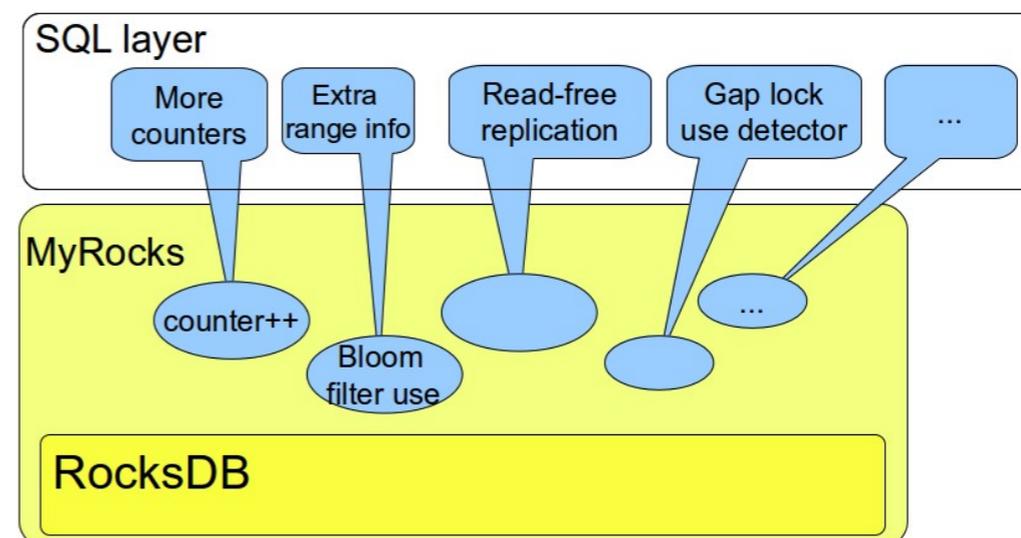


- Working

- General server functionality
- Optimizer-related
 - Bloom filter support
- Tests

- Non-working yet

- Interplay between the storage engine and the binlog
 - Group commit
 - Read-Free-Slave
 - GTID
- Gap Lock detector



diff -u mariarocks myrocks



3K line diff for the code

- CMake build changes
- MariaDB's differences in Extended Keys
- MariaDB's differences in Index Condition Pushdown
- MariaDB calls `ha_statistic_increment(...)` above the storage engine
- Implement `prepare_*_scan()`, Bloom Filter works for ORDER BY DESC
- Storage Engine API function signature changes
 - s/MySQL/MariaDB/ in names of structs, constants, etc
- #ifdef-away code related to replication and binlog
- #ifdef-away extra diagnostics like SHOW ENGINE TRANSACTION STATUS
- ...

8K lines diff for tests:

- MariaDB has different default values for SQL columns
- EXPLAIN output is slightly different
- Handler_xxx counters counted a bit differently
- MTR explicitly logs connection open/switch/etc
- `mysql-test/suite/$MYROCKS_TEST` → `storage/rocksdb/mysql-test/$MYROCKS_TEST`
- ...

Gap Lock Detector



- SQL level feature
- Detects queries that use gap locking
- Workflow
 - Run **with InnoDB**, detect and remove the querie
 - Migrate to MyRocks
- Objections to pushing
 - MyRocks has features on SQL layer!
- ?



- **MariaDB 10.2.5 RC2 includes MyRocks**
- Features
- **Packaging**
- Quality



- MyRocks is linked statically
- RocksDB is a submodule (@revno)
 - Not a dependency
- Can compile with compression libraries
 - Snappy
 - Zlib
 - ZStandard
 - ...



- MyRocks is built as **ha_rocksdm.so**
- RocksDB is a submodule (@revno)
 - Linked in statically in the .so
- Compression libraries
 - Bintar: links statically (snappy, zlib)
 - .deb/.rpm: package has a dependency (snappy, zlib)
 - Windows zip/MSI: zlib
- rocksdm_hotbackup is missing
 - but works (have an MDEV for this)



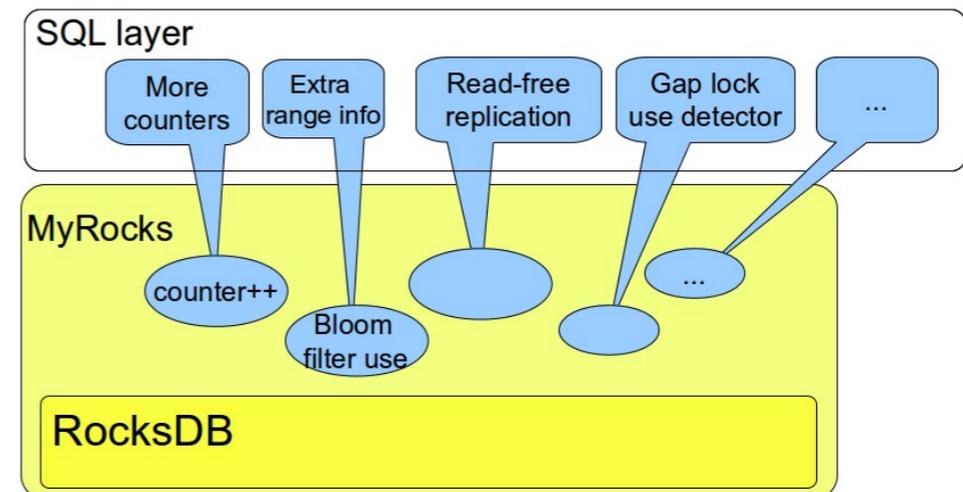
- MyRocks is built as **ha_rocksdm.so**
- MyRocks is compiled by default
 - Except for unsupported platforms
- MyRocks tests can be run as regular tests
 - `./mysql-test-run rocksdm.type_varchar`



- **MariaDB 10.2.5 RC2 includes MyRocks**
- Features
- Packaging
- **Quality**



- Alpha-maturity plugin in RC-level MariaDB Server
- MyRocks tests are run in MariaDB buildbot
- There are failing testcases
 - Not necessarily MyRocks bugs
 - Working on this
- There are missing features
 - Can finish those that touch only storage/rocksdb
 - May be harder if the feature touches the SQL layer.





Future



- Fix the failing tests
 - Some failures are from the upstream
- Usability improvements
- Interplay between the Storage Engine and Binlog
 - Get it to work
 - MariaDB's Group Commit
 - Per-engine **mysql.slave_gtid_pos**
 - Parallel Slave support?
- Keep merging from upstream.



Thanks!