MyRocks in MariaDB

Sergei Petrunia <sergey@mariadb.com>
MariaDB Shenzhen Meetup
November 2017

What is MyRocks

- #include <Yoshinori's talk>
- This talk is about MyRocks in MariaDB

MyRocks lives in Facebook's MySQL branch

- github.com/facebook/mysql-5.6
  - Will call this "FB/MySQL"
- MyRocks lives there in storage/rocksdb
- FB/MySQL is easy to use if you are Facebook
- Not so easy if you are not :-)

FB/mysql-5.6 – user perspective

- No binaries, no packages
  - Compile yourself from source
    - Dependencies, etc.
- No releases
  - (Is the latest git revision ok?)
- Has extra features
  - e.g. extra counters "confuse" monitoring tools.

FB/mysql-5.6 – dev perspective

- Targets a CentOS-type OS
  - Compiler, cmake version, etc.
  - Others may or may not [periodically] work
    - MariaDB/Percona file pull requests to fix
- Special command to compile
  - https://github.com/facebook/mysql-5.6/wiki/Build-Steps
- Special command to run tests
  - Test suite assumes a big machine
    - Some tests even a release build

Putting MyRocks in MariaDB

- Goals
  - Wider adoption
  - Ease of use
  - Ease of development
  - Have MyRocks in MariaDB
    - Use it with MariaDB features
- Means
  - Port MyRocks into MariaDB
  - Provide binaries and packages
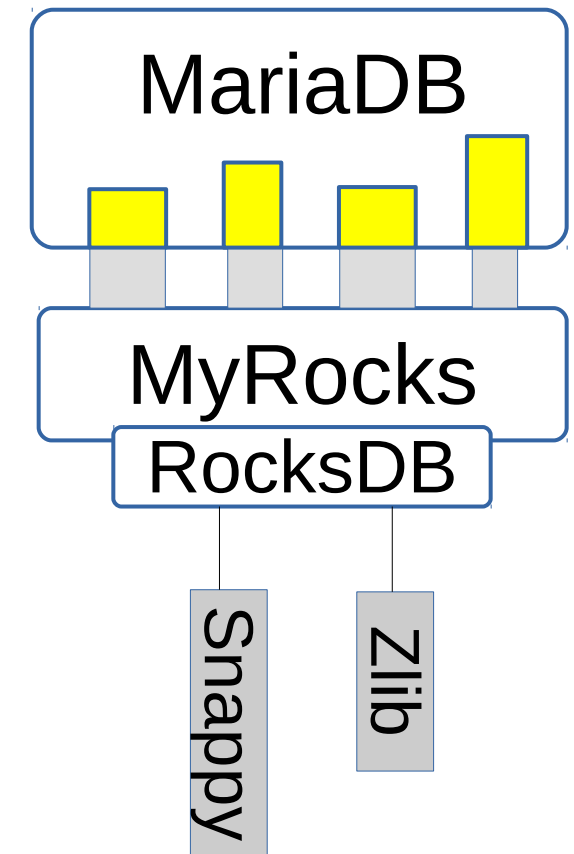
Status of
MyRocks in MariaDB

Status of MyRocks in MariaDB

- MariaDB 10.2 is GA (as of May, 2017)
- It includes an ALPHA version of MyRocks plugin
  - Working to improve maturity
- It's a loadable plugin (ha_rocksdb.so)
- Packages
  - Bintar, deb, rpm, win64 zip + MSI
  - deb/rpm have MyRocks .so and tools in a separate package.
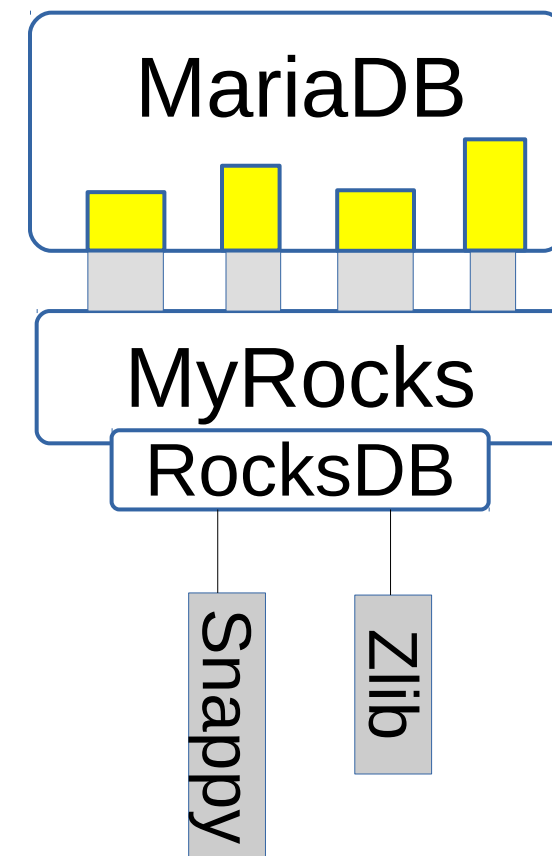
MariaDB®

Packaging for MyRocks in MariaDB

MyRocks and RocksDB library

- MyRocks is tied RocksDB@revno
  - RocksDB is a github submodule
  - No compatibility with other versions
- RocksDB is always compiled with MyRocks
- And linked-in statically
- Distros have a RocksDB package
  - Not using it.

MariaDB

MyRocks

RocksDB

Snappy

Zlib

Compression libraries

- RocksDB uses compression libraries
  - Snappy, Zlib, Bzip, LZ4, ZStandard
- Distros strongly prefer you to use OS packages
  - e.g. don't ship your own bzip
- MariaDB's MyRocks package has proper dependencies.

MyRocks is a loadable engine

- FB/MySQL-5.6 links MyRocks statically
- Considerations for MariaDB
  - Added size
  - Extra dependencies (compression libs)
  - Alpha-level maturity
- => MyRocks is packaged as loadable module
  - ha_rocksdb.so
  - Separate .deb/.rpm

MyRocks package

```
Package: mariadb-plugin-rocksdb
Source: mariadb-10.2
Version: 10.2.10+maria~trusty
Architecture: amd64
Maintainer: MariaDB Developers <maria-developers@lists.launchpad.net>
Installed-Size: 13047
Depends: mariadb-server-10.2, libc6 (>= 2.17), libsnappy1, libstdc++6 (>= 4.8.1),
zlib1g (>= 1:1.1.4)
Section: database
Priority: optional
Homepage: http://mariadb.org/
Description: RocksDB storage engine for MariaDB
 The RocksDB storage engine is a high performance storage engine, aimed
 at maximising storage efficiency while maintaining InnoDB-like performance.
 This package contains the RocksDB plugin for MariaDB.


-rw-r--r-- root/root    5162440 2017-10-30 17:40 ./usr/lib/mysql/plugin/ha_rocksdb.so
-rwxr-xr-x root/root    4066232 2017-10-30 17:40 ./usr/bin/sst_dump
-rwxr-xr-x root/root    4074456 2017-10-30 17:40 ./usr/bin/mysql_ldb
-rw-r--r-- root/root         40 2017-10-30 16:13 ./etc/mysql/mariadb.conf.d/rocksdb.cnf
```
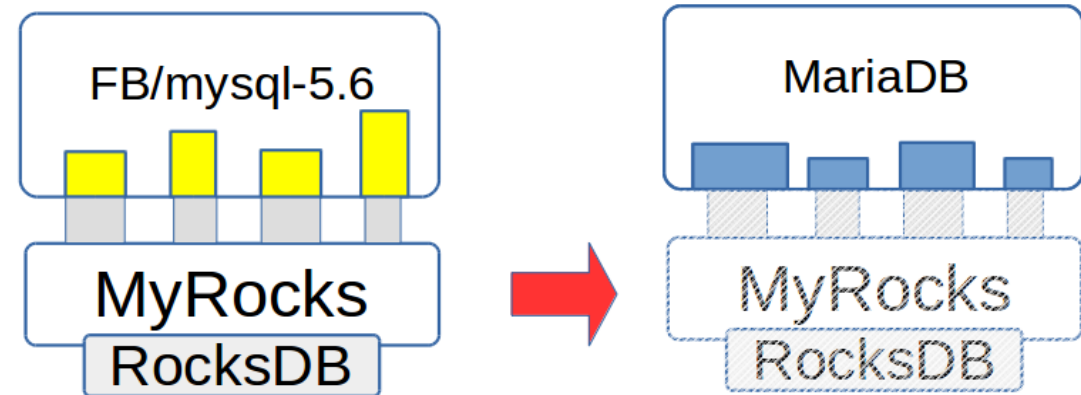
MariaDB®

MyRocks package status

- Available

  – Bintar, Deb, RPM, Win64 zip + MSI

- Compression libraries

  – Proper dependencies on practically important

  – Except for ZStandard

- Tools

  – myrocks_hotbackup is missing

    • This depends on getting replication to work

Integration of MyRocks into MariaDB

MyRocks integration

- MyRocks has interplay with SQL layer

- FB/5.6 [SQL layer] has extra features

- Need to un-couple from FB/5.6

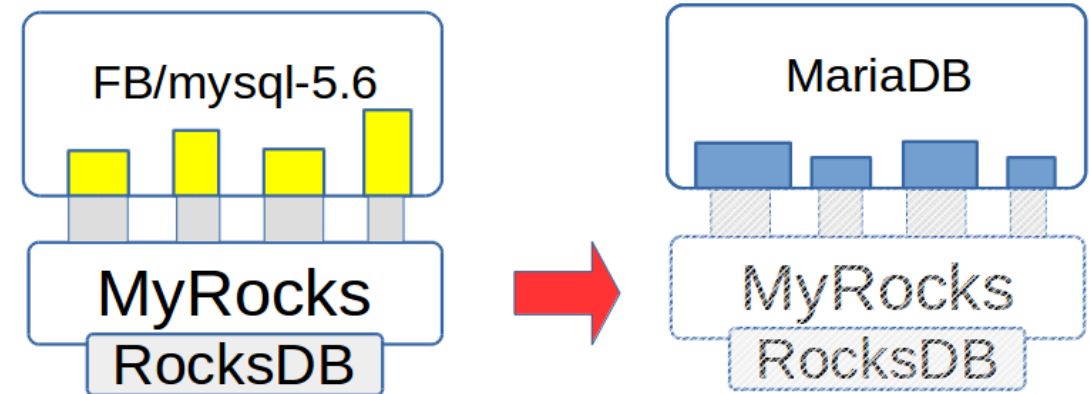- And couple with MariaDB's equivalents

MyRocks integration

- ## Most features
  - Generic SQL
  - Bloom filter
  - Crash-safe master
  - ...

- ## Not fully working yet:
  - Replication slave
  - Interplay with GTID

  - Backup is not packaged

FB/mysql-5.6

MyRocks
RocksDB

MariaDB

MyRocks
RocksDB

Keeping up to date with FB/MySQL

- MyRocks continues to evolve @ Facebook

- New changes are periodically merged into MariaDB

- "Merge tree" approach

  - Can view the merge status at
    https://github.com/MariaDB/mergetrees/commits/merge-myrocks

- Merging is still a manual process

  - But the amount of effort is reasonable

diff -u mysql/myrocks mariadb/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 ROCKSDB 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
- ...

MariaDB

MyRocks in MariaDB takeaways

- Currently ALPHA-maturity plugin in MariaDB 10.2
  - Most things work
  - GTID/Replication slave is work in progress
  - Working to improve maturity
- Packages and binaries are available
  - MyRocks is built as .so
- MyRocks keeps up to date with FB/MySQL.

MariaDB

Thanks!