MyRocks Troubleshooting

March, 30, 2017

Sveta Smirnova, Sergei Petrunia, Yoshinori Matsunobu, George O. Lorch III

Table of Contents

PERCONA

What you Need to Know About MyRocks?

Main Engine Characteristics

- Transactional
- Row-level locks
- LSM Tree

PERCONA

How Writes to LSM Performed

- Change request

PERCONA

How Writes to LSM Performed

- Change request
- MemTable

PERCONA

How Writes to LSM Performed

- Change request
- MemTable
- WAL

PERCONA

How Writes to LSM Performed

- Change request
- MemTable
- WAL
- Compaction

PERCONA

How Writes to LSM Performed

- Change request
- MemTable
- WAL
- Compaction
- SST
  - Immutable

PERCONA

Column Families

- Each index belongs to one column family

PERCONA

Column Families

- Each index belongs to one column family
- Column family can contain many indexes

PERCONA

Column Families

- Each index belongs to one column family
- Column family can contain many indexes
- Custom optimizations

PERCONA

Column Families

- Each index belongs to one column family
- Column family can contain many indexes
- Custom optimizations
- You can specify custom options for each family
  - rev:index_for_order_by_desc

PERCONA

Column Families

- Each index belongs to one column family
- Column family can contain many indexes
- Custom optimizations
- You can specify custom options for each family
  - rev:index_for_order_by_desc
- Do not create many column families
  - 20 is usually enough

PERCONA

Compaction Levels

- Level 0: up to 256MB

PERCONA

Compaction Levels

- Level 0: up to 256MB
- Level 1: up to 1G

PERCONA

Compaction Levels

- Level 0: up to 256MB
- Level 1: up to 1G
- Level 2: up to 10G

PERCONA

Compaction Levels

- Level 0: up to 256MB
- Level 1: up to 1G
- Level 2: up to 10G
- Level 3: up to 100G

Compaction Levels

- Level 0: up to 256MB
- Level 1: up to 1G
- Level 2: up to 10G
- Level 3: up to 100G
- Level 4: up to 1000G

Limitations

- Index-only access supported only for
  - BINARY
  - Collation latin1_bin
  - Collation utf8_bin

PERCONA

Limitations

- Index-only access supported only for
- No support for
  - Foreign Keys
  - Full Text Keys
  - Spatial Keys

PERCONA

Limitations

- Index-only access supported only for
- No support for
- Not recommended to use together with InnoDB
  - Mixing two transactional engines is dangerous!

PERCONA

How to Get MyRocks

- Percona
  - To build add Cmake option -DWITH_ROCKSDB=1
  - Binaries planned
- MariaDB
- Facebook MySQL Branch

PERCONA

How to Get MyRocks

- Percona
- MariaDB
  - Build instructions
  - Binaries planned
- Facebook MySQL Branch

PERCONA

How to Get MyRocks

- Percona
- MariaDB
- Facebook MySQL Branch
  - Facebook MySQL-5.6 branch
  - No binaries
  - You have to compile from sources

PERCONA

Data Corruption

Two Kinds of Corruptions

- Corrupted immutable files
  - Bad disk blocks
  - FS disaster
  - Not recoverable

PERCONA

Two Kinds of Corruptions

- Corrupted immutable files
- WAL file damaged
  - mysqld killed
  - OS restarted
  - Recoverable

PERCONA

Two Kinds of Corruptions

- Corrupted immutable files
- WAL file damaged
- In all cases
  - MySQL error log file
  - $datadir/.rocksdb/LOG
  - Enable core file
  - Send bug report

PERCONA

WAL Recovery: How it Works

- Controlled by rocksdb_wal_recovery_mode

PERCONA

WAL Recovery: How it Works

- Controlled by rocksdb_wal_recovery_mode
- 1: Fail to start, do not recover

PERCONA

WAL Recovery: How it Works

- Controlled by rocksdb_wal_recovery_mode
- 1: Fail to start, do not recover
- 0: If corrupted last entry: truncate and start

PERCONA

WAL Recovery: How it Works

- Controlled by rocksdb_wal_recovery_mode
- 1: Fail to start, do not recover
- 0: If corrupted last entry: truncate and start
- 2: Truncate everything after corrupted entry
  - Even not corrupted entries
  - Acceptable on slaves

PERCONA

WAL Recovery: How it Works

- Controlled by rocksdb_wal_recovery_mode
- 1: Fail to start, do not recover
- 0: If corrupted last entry: truncate and start
- 2: Truncate everything after corrupted entry
  - Even not corrupted entries
  - Acceptable on slaves
- 3: Truncate only corrupted entry
  - Most dangerous option

PERCONA

WAL Recovery: What can Be Lost

- 0: Last transaction

PERCONA

WAL Recovery: What can Be Lost

- 0: Last transaction
- 2: Last N trsansacitons

PERCONA

WAL Recovery: What can Be Lost

- 0: Last transaction
- 2: Last N trsansacitons
- 3: Corrupted transaction(s)
  - Can lead to total damage if following entries depend on missed transaction(s)

PERCONA

Inconsistent Data

Main Behavior Differences

- PostgreSQL-style snapshot model
  - Locking reads (UPDATE, SELECT ... FOR UPDATE) from snapshot
  - READ COMMITTED: at each statement
  - REPEATABLE READ: at the first statement of the transaction

PERCONA

Main Behavior Differences

- PostgreSQL-style snapshot model
- No Gap Lock support
  - Does not lock rows, scanned by SELECT
  - Does not lock rows which not exist
  - Does not lock ranges
  - Not compatible with pt-table-checksum

```
REPLACE INTO percona.checksums
SELECT * FROM app_tables WHERE key > X and key < Y;
```

  - Requires RBR for data consistency

PERCONA

Main Behavior Differences

- PostgreSQL-style snapshot model
- No Gap Lock support
  - With REPEATABLE READ Percona Server will return error when query requires gap lock

```
session 1> select * from employees.employees
        -> where emp_no > 20000 and emp_no < 21000 limit 10 for update;
ERROR 1105 (HY000): Using Gap Lock without full unique key in multi-table or
multi-statement transactions is not allowed. You need to either rewrite queries to use all
unique key columns in WHERE equal conditions,
or rewrite to single-table, single-statement transaction.
Query: select * from employees.employees
where emp_no > 20000 and emp_no < 21000 limit 10 for update
```

PERCONA

Main Behavior Differences

- PostgreSQL-style snapshot model
- No Gap Lock support
- Supported Transaction Isolation Levels
  - ~~READ UNCOMMITTED~~
  - READ COMMITTED
  - REPEATABLE READ
  - ~~SERIALIZABLE~~

PERCONA

Binary collations

- You can create table with any collation

```
session 1> create table ci_coll(f1 varchar(100)) engine=rocksdb;
Query OK, 0 rows affected (0.10 sec)

session 1> select COLUMN_NAME, CHARACTER_SET_NAME, COLLATION_NAME
        -> from information_schema.columns where table_name='ci_coll';
+-------------+--------------------+-------------------+
| COLUMN_NAME | CHARACTER_SET_NAME | COLLATION_NAME    |
+-------------+--------------------+-------------------+
| f1          | latin1             | latin1_swedish_ci |
+-------------+--------------------+-------------------+
1 row in set (0.01 sec)
```

PERCONA

Binary collations

- You can create table with any collation
- But could not create index

```
session 1> create index ci on ci_coll(f1);
ERROR 1105 (HY000): Unsupported collation on string indexed column test.ci_coll.f1
Use binary collation (binary, latin1_bin, utf8_bin).
```

PERCONA

Binary collations

- You can create table with any collation
- But could not create index
- Solution: `rocksdb_strict_collation_exceptions='TABLE_NAME'`

```
session 1> set global rocksdb_strict_collation_exceptions='ci_coll';
Query OK, 0 rows affected (0.00 sec)

session 1> create index ci on ci_coll(f1);
Query OK, 0 rows affected (0.22 sec)
Records: 0  Duplicates: 0  Warnings: 0
```

PERCONA

Backup

- mysqldump –single-transaction
- Facebook and MariaDB
  - Checks defaults-storage-engine
  - InnoDB?
    - Regular behavior
  - RocksDB?
    - `SET TRANSACTION ISOLATION LEVEL REPEATABLE READ`
    - `START TRANSACTION WITH CONSISTENT ROCKSDB SNAPSHOT`

PERCONA

Backup

- mysqldump –single-transaction
- Percona Server
  - rocksdb_skip_fill_cache exists?
    - MyRocks installed
    - Set it to 1
    - Enables rocksdb_bulk_load
  - `START TRANSACTION WITH CONSISTENT SNAPSHOT`
    - Compatible with MyRocks
    - Should be safe to backup together with InnoDB

PERCONA

Backup

- mysqldump –single-transaction
- myrocks_hotbackup

```
sveta@Thinkie:$ ~/build/mysql-5.6-fb/bin/myrocks_hotbackup -c ~/tmp/chkp
--user=root --password= -S var/tmp/mysqld.1.sock > backup.tar
2017-03-27 15:44:14.599 INFO Starting backup.
2017-03-27 15:44:14.605 INFO Set datadir: /home/sveta/build/ps-5.7/mysql-test/var/mysqld.1/data/
2017-03-27 15:44:14.605 INFO Creating checkpoint at /home/sveta/tmp/chkp/1
2017-03-27 15:44:14.817 INFO Created checkpoint at /home/sveta/tmp/chkp/1
2017-03-27 15:44:14.817 INFO Starting backup from snapshot: target files 4
...
2017-03-27 15:44:15.931 INFO MySQL misc backups done.
2017-03-27 15:44:15.931 INFO All Backups Done.
```

PERCONA

Backup

- mysqldump –single-transaction
- myrocks_hotbackup
  - Python script
  - Creates binary backups
  - Requires running server
  - Crash recovery happens at startup
  - Does not set any locks when copies frm files!
    - You must ensure there is no parallel DDL while backup is running

PERCONA

Locks

Limitations

- All uncommitted changes are in memory
  - Size of transaction limited

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 0
1 row in set (0.00 sec)
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 0
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 0
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 0
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
session1> UPDATE t SET f='foo' WHERE id=54321;
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 0
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
session1> UPDATE t SET f='foo' WHERE id=54321;
session2> UPDATE t SET f='foo' WHERE id=12345;
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 0
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
session1> UPDATE t SET f='foo' WHERE id=54321;
session2> UPDATE t SET f='foo' WHERE id=12345;
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 1
1 row in set (0.00 sec)
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 1
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 1
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 1
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
session1> UPDATE t SET f='foo' WHERE id=54321;
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 1
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
session1> UPDATE t SET f='foo' WHERE id=54321;
session2> UPDATE t SET f='foo' WHERE id=12345;
ERROR 1213 (40001): Deadlock found when trying to get lock; try restarting transaction
```

PERCONA

Limitations

- All uncommitted changes are in memory
- Deadlock detection OFF by default

```
information schema> select @@rocksdb_deadlock_detect\G
*************************** 1. row ***************************
@@rocksdb_deadlock_detect: 1
1 row in set (0.00 sec)
session1> START TRANSACTION;
session1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
session2> START TRANSACTION;
session2> SELECT * FROM t WHERE id=54321 FOR UPDATE;
session1> UPDATE t SET f='foo' WHERE id=54321;
session2> UPDATE t SET f='foo' WHERE id=12345;
ERROR 1213 (40001): Deadlock found when trying to get lock; try restarting transaction
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- InnoDB
  - SELECT ... FOR UPDATE

```
session 1> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 1> SELECT * FROM i WHERE id=12345 FOR UPDATE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.00 sec)
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- InnoDB
  - SELECT ... FOR UPDATE

```
session 2> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 2> SELECT * FROM i WHERE id=12345 FOR UPDATE;
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction
```

LOCK IN SHARE MODE

- Silently Ignored
- InnoDB
  - SELECT ... LOCK IN SHARE MODE

```
session 1> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 1> SELECT * FROM i WHERE id=12345 LOCK IN SHARE MODE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.01 sec)
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- InnoDB
  - SELECT ... LOCK IN SHARE MODE

```
session 2> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 2> SELECT * FROM i WHERE id=12345 LOCK IN SHARE MODE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.00 sec)
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- InnoDB
  - SELECT ... LOCK IN SHARE MODE

```
session 2> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 2> SELECT * FROM i WHERE id=12345 LOCK IN SHARE MODE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.00 sec)
session 2> update i set f='foo' where id=12345;
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction
```

LOCK IN SHARE MODE

- Silently Ignored
- MyRocks
  - SELECT ... FOR UPDATE

```
session 1> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 1> SELECT * FROM t WHERE id=12345 FOR UPDATE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.00 sec)
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- MyRocks
  - SELECT ... FOR UPDATE

```
session 2> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 2> SELECT * FROM i WHERE id=12345 FOR UPDATE;
ERROR 1205 (HY000): Lock wait timeout exceeded; try restarting transaction:
Timeout on index: test.t.PRIMARY
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- MyRocks
  - SELECT ... LOCK IN SHARE MODE

```
session 1> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 1> SELECT * FROM t WHERE id=12345 LOCK IN SHARE MODE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.01 sec)
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- MyRocks
  - SELECT ... LOCK IN SHARE MODE

```
session 2> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 2> SELECT * FROM t WHERE id=12345 LOCK IN SHARE MODE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.00 sec)
```

PERCONA

LOCK IN SHARE MODE

- Silently Ignored
- MyRocks
  - SELECT ... LOCK IN SHARE MODE

```
session 2> start transaction;
Query OK, 0 rows affected (0.00 sec)
session 2> SELECT * FROM t WHERE id=12345 LOCK IN SHARE MODE;
+-------+--------+
| id    | f      |
+-------+--------+
| 12345 | value1 |
+-------+--------+
1 row in set (0.00 sec)
session 2> update t set f='foo' where id=12345;
Query OK, 1 row affected (0.01 sec)
Rows matched: 1  Changed: 1  Warnings: 0
```

PERCONA

Tables in Information Schema

- ROCKSDB_TRX

```
mysql> select TRANSACTION_ID, STATE, WRITE_COUNT, LOCK_COUNT, TIMEOUT_SEC from rocksdb_trx\G
*************************** 1. row ***************************
         TRANSACTION_ID: 9
                  STATE: STARTED
            WRITE_COUNT: 3
             LOCK_COUNT: 2
            TIMEOUT_SEC: 1
WAITING_COLUMN_FAMILY_ID: 0
*************************** 2. row ***************************
         TRANSACTION_ID: 10
                  STATE: STARTED
            WRITE_COUNT: 0
             LOCK_COUNT: 1
            TIMEOUT_SEC: 1
2 rows in set (0.00 sec)
```

PERCONA

Tables in Information Schema

- ROCKSDB_TRX
- ROCKSDB_LOCKS

```
mysql> select * from rocksdb_locks\G
*************************** 1. row ***************************
COLUMN_FAMILY_ID: 0
  TRANSACTION_ID: 9
             KEY: 0000010180003039
            MODE: X
1 row in set (0.00 sec)
```

PERCONA

Tables in Information Schema

- ROCKSDB_TRX
- ROCKSDB_LOCKS
- No LOCK_WAITS table

PERCONA

SHOW ENGINE ROCKSDB TRANSACTION STATUS

```
mysql> show engine rocksdb transaction status\G
*************************** 1. row ***************************
  Type: SNAPSHOTS
  Name: rocksdb
Status:
============================================================
2017-03-28 03:32:51 ROCKSDB TRANSACTION MONITOR OUTPUT
============================================================
---------
SNAPSHOTS
---------
LIST OF SNAPSHOTS FOR EACH SESSION:
---SNAPSHOT, ACTIVE 60 sec
MySQL thread id 5, OS thread handle 0x7f6a3196f700, query id 138 localhost 127.0.0.1 root
lock count 1, write count 0
----------------------------------------
END OF ROCKSDB TRANSACTION MONITOR OUTPUT
========================================

1 row in set (0.00 sec)
```


PERCONA

SHOW ENGINE ROCKSDB TRANSACTION STATUS

- Facebook branch only
- Similar to TRANSACTIONS section in SHOW ENGINE INNODB STATUS
- Percona Server has same information in ROCKSDB_TRX

PERCONA

Slow Performance

Read Penalty

- Optimized for
  - Writes
  - Space savings

PERCONA

Read Penalty

- Optimized for
  - Writes
  - Space savings
- Reads
  - Have to traverse all compaction levels
  - L0-L2 are usually cached
  - Bloom filter

PERCONA

Optimizer Extensions

- Index Condition Pushdown supported

PERCONA

Optimizer Extensions

- Index Condition Pushdown supported
- MRR not supported

Optimizer Extensions

- Index Condition Pushdown supported
- MRR not supported
- No "index dives"
  - This is good!

PERCONA

Statistics for Small Tables

- Index statistics stored when MemTable flushed

PERCONA

Statistics for Small Tables

- Index statistics stored when MemTable flushed
- ANALYZE TABLE on small tables
  - Counter-intuitive

PERCONA

Statistics for Small Tables

- Index statistics stored when MemTable flushed
- ANALYZE TABLE on small tables
- Strange statistics even for PRIMARY KEYs

```
session 1> create table t1(id int not null auto_increment primary key) engine=rocksdb;
Query OK, 0 rows affected (0.11 sec)
session 1> insert into t1 values();
Query OK, 1 row affected (0.04 sec)
session 1> insert into t1 select null from t1;
Query OK, 1 row affected (0.03 sec)
Records: 1  Duplicates: 0  Warnings: 0
...
session 1> insert into t1 select null from t1;
Query OK, 1024 rows affected (0.11 sec)
Records: 1024  Duplicates: 0  Warnings: 0
```

PERCONA

Statistics for Small Tables

- Index statistics stored when MemTable flushed
- ANALYZE TABLE on small tables
- Strange statistics even for PRIMARY KEYs

```
session 1> show index from t1\G
*************************** 1. row ***************************
        Table: t1            |    Index_type: LSMTREE
   Non_unique: 0             |       Comment:
     Key_name: PRIMARY       | Index_comment:
 Seq_in_index: 1
  Column_name: id
    Collation: A
  Cardinality: 1914    -- Expected 2048
     Sub_part: NULL
       Packed: NULL
         Null:
                        1 row in set (0.00 sec)
```


PERCONA

Reverse Column Families

- Indexes optimized for single sorted order
  - Default: ascending

```
session 1> select * from joinit where g>30 and g < 32 order by s limit 1000, 1;
+--------+--------------------------------------+----------+----+
| i      | s                                    | t        | g  |
+--------+--------------------------------------+----------+----+
| 299721 | 07fe9309-13b2-11e7-a833-30b5c2208a0f | 15:28:12 | 31 |
+--------+--------------------------------------+----------+----+
1 row in set (1.34 sec)
session 1> select * from joinit where g>30 and g < 32 order by s desc limit 1000, 1;
+-------+--------------------------------------+----------+----+
| i     | s                                    | t        | g  |
+-------+--------------------------------------+----------+----+
| 82292 | f06c4f6c-13b1-11e7-a833-30b5c2208a0f | 15:27:40 | 31 |
+-------+--------------------------------------+----------+----+
1 row in set (1.58 sec)
```

Reverse Column Families

- Indexes optimized for single sorted order
- What if you need to run ORDER BY DESC?
  - Use reverse column families

```
session 1> alter table joinit drop index s;
Query OK, 0 rows affected (0.13 sec)
Records: 0  Duplicates: 0  Warnings: 0

session 1> alter table joinit add index s(s) comment 'rev:my_reversible_cf';
Query OK, 0 rows affected (14.72 sec)
Records: 0  Duplicates: 0  Warnings: 0
```

PERCONA

Reverse Column Families

- Indexes optimized for single sorted order
- What if you need to run ORDER BY DESC?

```
session 1> select * from joinit where g>30 and g < 32 order by s limit 1000, 1;
+--------+--------------------------------------+----------+----+
| i      | s                                    | t        | g  |
+--------+--------------------------------------+----------+----+
| 299721 | 07fe9309-13b2-11e7-a833-30b5c2208a0f | 15:28:12 | 31 |
+--------+--------------------------------------+----------+----+
1 row in set (1.58 sec)
session 1> select * from joinit where g>30 and g < 32 order by s desc limit 1000, 1;
+-------+--------------------------------------+----------+----+
| i     | s                                    | t        | g  |
+-------+--------------------------------------+----------+----+
| 82292 | f06c4f6c-13b1-11e7-a833-30b5c2208a0f | 15:27:40 | 31 |
+-------+--------------------------------------+----------+----+
1 row in set (1.42 sec)
```

Details on Reverse Column Families

```
mysql> SHOW ENGINE ROCKSDB STATUS\G
...
*************************** 5. row ***************************
  Type: CF_COMPACTION
  Name: rev:my_reversible_cf
Status:
** Compaction Stats [rev:my_reversible_cf] **
Level    Files   Size(MB) Score Read(GB)  Rn(GB) Rnp1(GB) Write(GB) Wnew(GB) Moved(GB) W-Amp ...
----------------------------------------------------------------------------------------------------
 Sum      1/0       61.51   0.0      0.0     0.0      0.0       0.1       0.1       0.0   1.0  ...
 Int      0/0        0.00   0.0      0.0     0.0      0.0       0.1       0.1       0.0   1.0  ...
Uptime(secs): 215.1 total, 215.1 interval
Flush(GB): cumulative 0.000, interval 0.000
AddFile(GB): cumulative 0.060, interval 0.060
AddFile(Total Files): cumulative 1, interval 1
AddFile(L0 Files): cumulative 0, interval 0
AddFile(Keys): cumulative 1048576, interval 1048576
...
```

Bloom Filter

- Checks if key may exist or not without reading data

PERCONA

Bloom Filter

- Checks if key may exist or not without reading data
- If not exists: skip read i/o

PERCONA

Bloom Filter

- Checks if key may exist or not without reading data
- If not exists: skip read i/o

```
rocksdb_default_cf_options=rocksdb_default_cf_options=prefix_extractor=capped:12;
block_based_table_factory={cache_index_and_filter_blocks=1;filter_policy=bloomfilter:10:false;
whole_key_filtering=0}
```

PERCONA

Bloom Filter

- Checks if key may exist or not without reading data
- If not exists: skip read i/o

```
rocksdb_default_cf_options=rocksdb_default_cf_options=prefix_extractor=capped:12;
block_based_table_factory={cache_index_and_filter_blocks=1;filter_policy=bloomfilter:10:false;
whole_key_filtering=0}
```

- Condition must be larger than or equal

PERCONA

Bloom Filter

- Checks if key may exist or not without reading data
- If not exists: skip read i/o

```
rocksdb_default_cf_options=rocksdb_default_cf_options=prefix_extractor=capped:12;
block_based_table_factory={cache_index_and_filter_blocks=1;filter_policy=bloomfilter:10:false;
whole_key_filtering=0}
```

- Condition must be larger than or equal
- Can be set per column family

PERCONA

Bloom Filter

- Checks if key may exist or not without reading data
- If not exists: skip read i/o

```
rocksdb_default_cf_options=rocksdb_default_cf_options=prefix_extractor=capped:12;
block_based_table_factory={cache_index_and_filter_blocks=1;filter_policy=bloomfilter:10:false;
whole_key_filtering=0}
```

- Condition must be larger than or equal
- Can be set per column family
- Stored in SST files

PERCONA

Bulk Operations Options

- MyRocks designed for small transactions

PERCONA

Bulk Operations Options

- MyRocks designed for small transactions
- There are optimizations for bulk operations

Bulk Operations Options

- MyRocks designed for small transactions
- There are optimizations for bulk operations
- Data Loading
  - Drop secondary keys
  - rocksdb_bulk_load=1
    - Data must be inserted in Primary Key order
    - None of the data may overlap with existing
    - Data will not be visible until load finishes

PERCONA

Bulk Operations Options

- MyRocks designed for small transactions
- There are optimizations for bulk operations
- Massive DELETEs
  - DELETE ... LIMIT ... are not effective
  - rocksdb-master-skip-tx-api
    - Disables Transaction API
    - Enables WriteBatch API
    - UPDATE and DELETEs are faster
    - There is no row lock
    - You must ensure no concurrent operation running!

PERCONA

Bulk Operations Options

- MyRocks designed for small transactions
- There are optimizations for bulk operations
- Massive DELETEs
  - DELETE ... LIMIT ... are not effective
  - Blind DELETEs by PK
    - rocksdb-blind-delete-primary-key
    - DELETES by Primary Key
    - Works:

      ```
      DELETE FROM t WHERE id IN (1, 2, 3, 4, 5, 6, ...., 10000)
      ```

    - Does not work:

      ```
      DELETE .. WHERE id < 10
      ```

PERCONA

Bulk Operations Options

- MyRocks designed for small transactions
- There are optimizations for bulk operations
- Massive DELETEs
  - DELETE ... LIMIT ... are not effective
  - Implicit commits on large DELETEs
    - rocksdb-commit-in-the-middle
    - Commit each rocksdb_bulk_load_size rows
    - On failure: no way to find out how much rows were deleted

PERCONA

Instruments

SHOW ENGINE ROCKSDB STATUS

- DBSTATS
  - Uptime

```
session 1> SHOW ENGINE ROCKSDB STATUS\G
*************************** 1. row ***************************
  Type: DBSTATS
  Name: rocksdb
Status:
** DB Stats **
Uptime(secs): 4138.8 total, 3.3 interval
```

PERCONA

SHOW ENGINE ROCKSDB STATUS

- DBSTATS
  - IO statistics

```
Cumulative writes: 8 writes, 10 keys, 8 commit groups, 0.9 writes per commit group,
ingest: 0.00 GB, 0.00 MB/s
Cumulative WAL: 8 writes, 8 syncs, 0.89 writes per sync, written: 0.00 GB, 0.00 MB/s
Cumulative stall: 00:00:0.000 H:M:S, 0.0 percent
Interval writes: 0 writes, 0 keys, 0 commit groups, 0.0 writes per commit group,
ingest: 0.00 MB, 0.00 MB/s
Interval WAL: 0 writes, 0 syncs, 0.00 writes per sync, written: 0.00 MB, 0.00 MB/s
Interval stall: 00:00:0.000 H:M:S, 0.0 percent
```

PERCONA

SHOW ENGINE ROCKSDB STATUS

- DBSTATS
  - Level 0-4 latency histograms

```
** Level 0 read latency histogram (micros):
Count: 17 Average: 40.0000  StdDev: 94.04
Min: 4  Median: 12.2000  Max: 403
Percentiles: P50: 12.20 P75: 13.90 P99: 403.00 P99.9: 403.00 P99.99: 403.00
------------------------------------------------------
[       3,       4 )        3  17.647%  17.647% ####
[       5,       6 )        1   5.882%  23.529% #
[       7,       8 )        2  11.765%  35.294% ##
[       9,      10 )        1   5.882%  41.176% #
[      10,      12 )        1   5.882%  47.059% #
[      12,      14 )        5  29.412%  76.471% ######
[      18,      20 )        2  11.765%  88.235% ##
[     100,     120 )        1   5.882%  94.118% #
[     400,     450 )        1   5.882% 100.000% #
```

SHOW ENGINE ROCKSDB STATUS

- DBSTATS
- CF_COMPACTION
  - Compaction statistics per family
    - default
    - __system__
    - custom

PERCONA

SHOW ENGINE ROCKSDB STATUS

- DBSTATS
- CF_COMPACTION
- Memory_Stats

```
*************************** 5. row ***************************
  Type: Memory_Stats
  Name: rocksdb
Status:
MemTable Total: 2624
MemTable Unflushed: 2624
Table Readers Total: 0
Cache Total: 811101
Default Cache Capacity: 0
5 rows in set (0.00 sec)
```

PERCONA

ENGINE ROCKSDB TRANSACTION STATUS

- Facebook only

```
mysql> show engine rocksdb transaction status\G
*************************** 1. row ***************************
  Type: SNAPSHOTS
  Name: rocksdb
Status:
============================================================
2017-03-28 04:03:00 ROCKSDB TRANSACTION MONITOR OUTPUT
============================================================
...
----------------------------------------
END OF ROCKSDB TRANSACTION MONITOR OUTPUT
========================================

1 row in set (0.00 sec)
```

ENGINE ROCKSDB TRANSACTION STATUS

- Facebook only
- List of running transactions

```
---------
SNAPSHOTS
---------
LIST OF SNAPSHOTS FOR EACH SESSION:
---SNAPSHOT, ACTIVE 187 sec
MySQL thread id 9, OS thread handle 0x7f6a2ea32700, query id 541 localhost 127.0.0.1 root
lock count 1, write count 0
---SNAPSHOT, ACTIVE 189 sec
MySQL thread id 10, OS thread handle 0x7f6a2e9e1700, query id 544 localhost 127.0.0.1 root
User sleep
SELECT sleep(10) FROM t WHERE id=12345 FOR UPDATE
lock count 3, write count 0
```

PERCONA

GLOBAL STATUS

- Rows statistics
  - deleted
  - inserted
  - read
  - updated

PERCONA

GLOBAL STATUS

- Rows statistics
- Counters
  - rocksdb_number_block_not_compressed
  - rocksdb_number_deletes_filtered
  - rocksdb_number_keys_read
  - ...

PERCONA

GLOBAL STATUS

- Rows statistics
- Counters
- Block cache

PERCONA

GLOBAL STATUS

- Rows statistics
- Counters
- Block cache
- Bloom filter

PERCONA

GLOBAL STATUS

- Rows statistics
- Counters
- Block cache
- Bloom filter
- IO
  - rocksdb_bytes_read—written
  - rocksdb_compact_read—write_bytes
  - rocksdb_flush_write_bytes
  - rocksdb_write_*

GLOBAL STATUS

- Rows statistics
- Counters
- Block cache
- Bloom filter
- IO
- Compaction

PERCONA

GLOBAL STATUS

- Rows statistics
- Counters
- Block cache
- Bloom filter
- IO
- Compaction
- Memtable

PERCONA

GLOBAL STATUS

- Rows statistics
- Counters
- Block cache
- Bloom filter
- IO
- Compaction
- Memtable
- rocksdb_snapshot_conflict_errors

PERCONA

sst_dump

- Checks, scans and dumps SST files

PERCONA

sst_dump

- Checks, scans and dumps SST files
- Check

```
sst_dump --file=var/mysqld.1/data/.rocksdb/000012.sst --command=check --verify_checksum
from [] to []
Process var/mysqld.1/data/.rocksdb/000012.sst
Sst file format: block-based
```

PERCONA

sst_dump

- Checks, scans and dumps SST files
- Scan

```
sst_dump --file=var/mysqld.1/data/.rocksdb/000012.sst --command=scan | head
from [] to []
Process var/mysqld.1/data/.rocksdb/000012.sst
Sst file format: block-based
' => "C?GeorgiFacell???
' => ?XBezalel?Simmel???
' => ?OPartoBamford???
' => ?D ChirstianKoblick???
' => 5FKyoichMaliniak???
' => ?B?AnnekePreusig??
' => ?JTzvetan Zielinski???
```

PERCONA

sst_dump

- Checks, scans and dumps SST files
- Scan

```
sst_dump --file=var/mysqld.1/data/.rocksdb/000012.sst --command=scan --output_hex | head
from [] to []
Process var/mysqld.1/data/.rocksdb/000012.sst
Sst file format: block-based
'0000010080002711' @ 40: 1 => 22430F0647656F726769074661636566C6C6F01DA840F
'0000010080002712' @ 41: 1 => C2580F0742657A616C656C0653696D6D656C0275830F
'0000010080002713' @ 42: 1 => 834F0F05506172746F0742616D666F7264011C850F
'0000010080002714' @ 43: 1 => A1440F0943686972737469616E074B6F626C69636B0181850F
'0000010080002715' @ 44: 1 => 35460F074B796F69636869084D616C696E69616B012C8B0F
'0000010080002716' @ 45: 1 => 94420F06416E6E656B650750726575736967 02C28A0F
'0000010080002717' @ 46: 1 => B74A0F07547A766574616E095A69656C696E736B69024A8A0F
```

sst_dump

- Checks, scans and dumps SST files
- Raw

```
sst_dump --file=var/mysqld.1/data/.rocksdb/000012.sst --command=rawfrom [] to []
Process var/mysqld.1/data/.rocksdb/000012.sst
Sst file format: block-based
raw dump written to file var/mysqld.1/data/.rocksdb/000012_dump.txt
```

PERCONA

sst_dump

- Checks, scans and dumps SST files
- Raw

```
Footer Details:
--------------------------------------
  checksum: 1
  metaindex handle: 92D1801323
  index handle: BAD18013AD8816
  footer version: 2
  table_magic_number: 9863518390377041911
```

PERCONA

sst_dump

- Checks, scans and dumps SST files
- Raw

```
Metaindex Details:
--------------------------------------
  Properties block handle: 8FC68013FE0A

Table Properties:
--------------------------------------
  # data blocks: 9781
  # entries: 1330253
  raw key size: 30820386
  raw average key size: 23.168815
  raw value size: 14034718
  raw average value size: 10.550413
  data block size: 39854863
  index block size: 361522
  ...
```


PERCONA

sst_dump

- Checks, scans and dumps SST files
- Raw

```
Index Details:
--------------------------------------
  Block key hex dump: Data block handle
  Block key ascii

  HEX    0000010080002784: 00F31F
  ASCII
  ------
  HEX    00000100800027F6: F81FDB1F
  ASCII
...
  ------
  HEX    000001068000902A0F8CDB: DEC40000DB8E0F
  ASCII  : <DE> <C4>
  ------
```


PERCONA

sst_dump

- Checks, scans and dumps SST files
- Raw

```
Data Block Summary:
--------------------------------------
  # data blocks: 9781
  min data block size: 536
  max data block size: 4091
  avg data block size: 4069.722728
```

PERCONA

mysql_ldb

- LevelDB Tool
  - Manages, backups, modify LevelDB files

PERCONA

mysql_ldb

- LevelDB Tool
  - Manages, backups, modify LevelDB files
- Check for consistency

```
sveta@Thinkie:$ ../bin/mysql_ldb checkconsistency --db=var/mysqld.1/data/.rocksdb
OK
```

PERCONA

mysql_ldb

- LevelDB Tool
  - Manages, backups, modify LevelDB files
- Check for consistency

```
sveta@Thinkie:$ ../bin/mysql_ldb checkconsistency --db=var/mysqld.1/data/.rocksdb
OK
```

- Dump WAL files

```
sveta@Thinkie:$ ../bin/mysql_ldb dump_wal --walfile=var/mysqld.1/data/.rocksdb/000096.log
6016356,0,12,0,
```

PERCONA

Information Schema

- ROCKSDB_GLOBAL_INFO
  - Global stats

```
mysql> select * from ROCKSDB_GLOBAL_INFO;
+--------------+--------------+----------------+
| TYPE         | NAME         | VALUE          |
+--------------+--------------+----------------+
| MAX_INDEX_ID | MAX_INDEX_ID | 262            |
| CF_FLAGS     | 0            | default [0]    |
| CF_FLAGS     | 1            | __system__ [0] |
+--------------+--------------+----------------+
3 rows in set (0.00 sec)
```

Information Schema

- ROCKSDB_CFSTATS
  - Column family statistics

```
mysql> select * from ROCKSDB_CFSTATS;
+-------------+-------------------------------+----------+
| CF_NAME     | STAT_TYPE                     | VALUE    |
+-------------+-------------------------------+----------+
| __system__  | NUM_IMMUTABLE_MEM_TABLE       |        0 |
| __system__  | MEM_TABLE_FLUSH_PENDING       |        0 |
| __system__  | COMPACTION_PENDING            |        0 |
| __system__  | CUR_SIZE_ACTIVE_MEM_TABLE     |     2736 |
| __system__  | CUR_SIZE_ALL_MEM_TABLES       |     2736 |
| __system__  | NUM_ENTRIES_ACTIVE_MEM_TABLE  |       29 |
...
| default     | CUR_SIZE_ACTIVE_MEM_TABLE     | 44997040 |
| default     | CUR_SIZE_ALL_MEM_TABLES       | 44997040 |
| default     | NUM_ENTRIES_ACTIVE_MEM_TABLE  |  1054330 |
...
```

Information Schema

- ROCKSDB_TRX
  - Currently running transactions
  - Filled when MyRocks table accessed

```
mysql> select * from ROCKSDB_TRX\G
*************************** 1. row ***************************
          TRANSACTION_ID: 337
                   STATE: STARTED
                    NAME:            |           SKIP_TRX_API: 0
             WRITE_COUNT: 0          |              READ_ONLY: 0
              LOCK_COUNT: 0          | HAS_DEADLOCK_DETECTION: 0
             TIMEOUT_SEC: 1          |   NUM_ONGOING_BULKLOAD: 0
             WAITING_KEY:            |               THREAD_ID: 6394
WAITING_COLUMN_FAMILY_ID: 0          |                   QUERY:
          IS_REPLICATION: 0
```

Information Schema

- ROCKSDB_CF_OPTIONS
  - Column family options

```
mysql> select * from ROCKSDB_CF_OPTIONS where CF_NAME='my_cf'\G
*************************** 1. row ***************************
    CF_NAME: my_cf
OPTION_TYPE: COMPARATOR
      VALUE: RocksDB_SE_v3.10
*************************** 2. row ***************************
    CF_NAME: my_cf
OPTION_TYPE: MERGE_OPERATOR
      VALUE: NULL
...
```

Information Schema

- ROCKSDB_COMPACTION_STATS
  - Compaction statistics

```
mysql> select * from ROCKSDB_COMPACTION_STATS where value > 0;
+---------+-------+------------+-------+
| CF_NAME | LEVEL | TYPE       | VALUE |
+---------+-------+------------+-------+
| default | L0    | AvgSec     |     4 |
| default | L0    | CompCount  |     2 |
| default | L0    | CompSec    |     9 |
...
| my_cf   | L6    | SizeMB     |     5 |
| my_cf   | L6    | WriteMBps  |   239 |
| my_cf   | Sum   | CompCount  |     1 |
| my_cf   | Sum   | NumFiles   |     1 |
| my_cf   | Sum   | SizeMB     |     5 |
| my_cf   | Sum   | WriteMBps  |   239 |
+---------+-------+------------+-------+
20 rows in set (0.00 sec)
```


PERCONA

Information Schema

- ROCKSDB_DBSTATS
  - Engine statistics

```
mysql> select * from ROCKSDB_DBSTATS;
+-----------------------+----------+
| STAT_TYPE             | VALUE    |
+-----------------------+----------+
| DB_BACKGROUND_ERRORS  |        0 |
| DB_NUM_SNAPSHOTS      |        0 |
| DB_OLDEST_SNAPSHOT_TIME |        0 |
| DB_BLOCK_CACHE_USAGE  | 11222831 |
+-----------------------+----------+
4 rows in set (0.01 sec)
```

PERCONA

Information Schema

- ROCKSDB_DDL
  - MyRocks Tables

```
mysql> select * from ROCKSDB_DDL\G
*************************** 1. row ***************************
     TABLE_SCHEMA: employees
       TABLE_NAME: dept_emp
   PARTITION_NAME: NULL
       INDEX_NAME: PRIMARY
    COLUMN_FAMILY: 0
     INDEX_NUMBER: 260
       INDEX_TYPE: 1
KV_FORMAT_VERSION: 11
               CF: default
..
```

PERCONA

Information Schema

- ROCKSDB_INDEX_FILE_MAP
  - SST files statistics

```
mysql> select * from ROCKSDB_INDEX_FILE_MAP\G
*************************** 1. row ***************************
      COLUMN_FAMILY: 0
       INDEX_NUMBER: 256
           SST_NAME: 000012.sst
           NUM_ROWS: 300024
          DATA_SIZE: 9111608
      ENTRY_DELETES: 0
ENTRY_SINGLEDELETES: 0
       ENTRY_MERGES: 0
       ENTRY_OTHERS: 0
...
```

PERCONA

Information Schema

- ROCKSDB_LOCKS
  - Currently acquired locks

```
mysql> select * from ROCKSDB_LOCKS;
+-------------------+----------------+-------------------+------+
| COLUMN_FAMILY_ID | TRANSACTION_ID | KEY               | MODE |
+-------------------+----------------+-------------------+------+
|                 0 |            339 | 0000010080004e20 | X    |
|                 0 |            339 | 0000010080005208 | X    |
+-------------------+----------------+-------------------+------+
2 rows in set (0.00 sec)
```

PERCONA

Information Schema

- ROCKSDB_PERF_CONTEXT
  - Session performance

```
session 1> select * from information_schema.ROCKSDB_PERF_CONTEXT\G
*************************** 1. row ***************************
  TABLE_SCHEMA: employees
    TABLE_NAME: dept_emp
PARTITION_NAME: NULL
     STAT_TYPE: USER_KEY_COMPARISON_COUNT
         VALUE: 0
*************************** 2. row ***************************
  TABLE_SCHEMA: employees
    TABLE_NAME: dept_emp
PARTITION_NAME: NULL
     STAT_TYPE: BLOCK_CACHE_HIT_COUNT
         VALUE: 0
...
```

PERCONA

Information Schema

- ROCKSDB_PERF_CONTEXT_GLOBAL
  - Global performance

```
mysql> select * from ROCKSDB_PERF_CONTEXT_GLOBAL;
+--------------------------------+-------+
| STAT_TYPE                      | VALUE |
+--------------------------------+-------+
| USER_KEY_COMPARISON_COUNT      |     0 |
| BLOCK_CACHE_HIT_COUNT          |     0 |
| BLOCK_READ_COUNT               |     0 |
| BLOCK_READ_BYTE                |     0 |
| BLOCK_READ_TIME                |     0 |
| BLOCK_CHECKSUM_TIME            |     0 |
...
```

PERCONA

.rocksdb/LOG

- MyRocks debugging information

```
2017/03/24-15:04:37.517806 7f920676a740 RocksDB version: 5.0.0
2017/03/24-15:04:37.517971 7f920676a740
Git sha rocksdb_build_git_sha:bc5d7b70299b763127f3714055a63ebe7e04ad47
2017/03/24-15:04:37.517973 7f920676a740 Compile date Mar 21 2017
2017/03/24-15:04:37.517975 7f920676a740 DB SUMMARY
2017/03/24-15:04:37.518066 7f920676a740 SST files in ./.rocksdb dir, Total Num: 0, files:
2017/03/24-15:04:37.518069 7f920676a740 Write Ahead Log file in ./.rocksdb:
2017/03/24-15:04:37.518071 7f920676a740                         Options.error_if_exists: 0
2017/03/24-15:04:37.518084 7f920676a740                       Options.create_if_missing: 1
...
```

PERCONA

Specific Debugging Options

- rocksdb_info_log_level=debug_level
  - Sets the log level for printing error messages
  - Available values
    - debug_level
    - info_level

```
2017/03/27-18:58:25.225140 7fa43dffe700 [default] [JOB 4435]
Generated table #2947184: 76 keys, 2024 bytes
2017/03/27-18:58:25.225163 7fa43dffe700 EVENT_LOG_v1 {"time_micros": 1490666305225148,
"cf_name": "default", "job": 4435, "event": "table_file_creation",
"file_number": 2947184, "file_size": 2024, "table_properties": {"data_size": 831,
"index_size": 43, "filter_size": 69, "raw_key_size": 1824, "raw_average_key_size": 24,
"raw_value_size": 0, "raw_average_value_size": 0, "num_data_blocks": 1,
"num_entries": 76, "filter_policy_name": "rocksdb.BuiltinBloomFilter",
"__indexstats__": "[...1  records...]", "kDeletedKeys": "0",
"kMergeOperands": "0"}}
2017/03/27-18:58:25.225215 7fa43dffe700 [default] [JOB 4435]
Compacted 1@4 + 30@5 files to L5 => 862898990 bytes
```

Specific Debugging Options

- rocksdb_info_log_level=debug_level
  - Sets the log level for printing error messages
  - Available values
    - debug_level
    - info_level
    - warn_level
    - error_level - default
    - fatal_level
  - Written to
    - $datadir/.rocksdb/LOG
    - MySQL error log file

PERCONA

Specific Debugging Options

- rocksdb_info_log_level=debug_level
- rocksdb_perf_context_level
  - Sets the level for perf context plugins

```
information schema> set global rocksdb_perf_context_level = 2;
information schema> select * from rocksdb_perf_context_global where value != 0;
+---------------------------+---------+
| STAT_TYPE                 | VALUE   |
+---------------------------+---------+
| USER_KEY_COMPARISON_COUNT | 5105904 |
| BLOCK_CACHE_HIT_COUNT     |  183831 |
| INTERNAL_KEY_SKIPPED_COUNT |  181119 |
| SEEK_ON_MEMTABLE_COUNT    |       3 |
| SEEK_CHILD_SEEK_COUNT     |       6 |
| IO_THREAD_POOL_ID         |      12 |
+---------------------------+---------+
6 rows in set (0.00 sec)
```


PERCONA

Summary

MyRocks Troubleshooting: Summary

- Space-optimized transactional engine
  - Fast writes
  - Great compression
  - Read optimizations
- Main differences from InnoDB
  - Transaction snapshots
  - Locking
  - IO
- Rich set of diagnostic tools

PERCONA

More information

- MyRocks Deep Dive
- MyRocks Wiki
- MyRocks @MariaDB
- MyRocks Bugs @Percona
- MyRocks Bugs @Facebook
- Percona Server GitHub
- Facebook MySQL 5.6 Tree

Time for questions

???

PERCONA

Thank you!

http://www.slideshare.net/SvetaSmirnova

https://twitter.com/svetsmirnova

PERCONA

DATABASE PERFORMANCE
MATTERS