Introduction To MariaDB

--现状和发展中的生态系统

Greeting From Monty

Topics

1) What is MariaDB
2) What new in the different MariaDB releases
3) Some benchmarks
4) MariaDB and NoSQL
5) Why next release is called MariaDB 10.0
6) Is MariaDB and MySQL future proof ?
7) Some announcements...
8) Conclusions

Why MariaDB was created

"Save the People, Save the Product"

- To keep the MySQL talent together
- To ensure that a free version of MySQL always exists
- To get one community developed and maintained branch
- Work with other MySQL forks/branches to share knowhow and code

What is MariaDB

- Open Source, binary compatible, superset of MySQL:
- True drop in replacement of MySQL
  - Same data on disk and on the wire.
  - Same file names, same sockets and ports.
- Created and maintained by the same people that created MySQL.
- Open development: Developed together with the community.
- True fork, not just a patch set upon MySQL.
  - MariaDB is not depending on MySQL for future development.
- More plugins, more features, faster, better code quality.
  - 30 man years more development than MySQL
- GPL-only server license (no closed source extensions)

MariaDB server releases

- MariaDB 5.1 was released as stable in February 2010
- MariaDB 5.2 was released as stable in November 2010
- MariaDB 5.3 was released as stable in April 2012
- MariaDB 5.5 was released as stable in April 2012
- MariaDB 10.0 was released as alpha in November 2012
  - Goal is to be released as stable in April

- MariaDB-Galera (multi-master) was released as stable in December 2012 after a lot of testing.

The MariaDB releases

- MariaDB 5.1 (based on MySQL 5.1)
  - Better build & test system, code cleanups, community patches, new storage engines, table elimination.
- MariaDB 5.2 (based on MariaDB 5.1)
  - Community features that did not go into 5.1:
    - Virtual columns
    - Extended User Statistics
    - Segmented MyISAM key cache (faster multi user!)
- MariaDB 5.3 (based on MariaDB 5.2)
  - Optimizer features (faster subquerier, joins etc)
  - Microsecond, dynamic columns, faster HANDLER etc.
  - Better replication (group commit, more options)
- MariaDB 5.5 (based on MariaDB 5.3 and MySQL 5.5)

MariaDB 5.3 and NoSQL

The main reasons for using NoSQL are:

- Handling of unstructured data (not everything is table and fixed number of columns)
- Faster replication (usually with 'unconventional' shortcuts)

- The same way MySQL with it's storage engine interface can handle both transactional and datawarehousing , we are extending MariaDB to be a bridge between SQL and NoSQL.

MariaDB as a bridge between SQL and NoSQL

- Up to 50 % faster HANDLER commands (HANDLER READ …)
  - Up to 530,000 queres/second measured(*)
- HandlerSocket compiled in (Direct access to InnoDB)
- Dynamic columns (each row can have different set of columns)
  - Building block for NoSQL storage engines
- Storage engine for Cassandra
  - You can read, write, update and join with Cassandra
- We are working on a storage engine for LevelDB

(*) Stephane Varoqui's blog:
http://varokism.blogspot.com/2011/01/20-to-50-improvement-in-mariadb-53.html

What's new in MariaDB 5.5

- Significantly more efficient thread pool
- Non-blocking client API Library (MWL#192)
- SphinxSE updated to version 2.0.4.
- Extended Keys support for XtraDB and InnoDB
- New LIMIT ROWS EXAMINED optimization.
  - Limits max number rows examined for a query
- Lots of security fixes, new status variables and small enhancements.

- https://kb.askmonty.org/en/mariadb-vs-mysql-features/
- https://kb.askmonty.org/en/what-is-mariadb-55/

MariaDB 10.0

MariaDB 10.0 is MariaDB 5.5 + some features from MySQL 5.6 + some new features

Features back ported from MySQL 5.6:

- All InnoDB changes (done)
- Performance schema changes (done)
- Read only transaction (significant InnoDB optimization) (done)
- Online ALTER TABLE (in progress)

Features from MySQL 5.6 that are reimplemented:

- Better error message (with system error string) (done)
- NOW() as default value for datetime (done)
- Global transaction ID for replication (in progress)
- Parallel replication (much better implementation)

MariaDB 10.0

New features:

- SHOW EXPLAIN (see what other thread is doing) (done)
- Multi source (one slave can have many masters) (done)
- Faster ALTER TABLE with UNIQUE index (done)
- DELETE … RETURNING (in review)
- Even faster group commit (in progress)
- Storage engine for Cassandra (done)
- Storage engine for Leveldb (in progress)
- Per thread memory usage (done)

For full list, see http://kb.askmonty.org/v/plans-for-10x

Optimizations comparison

Features	MariaDB 5.3/5.5	MySQL 5.5	MySQL 5.6
Index Condition Pushdown (ICP)	Yes		Yes
Disk-sweep Multi-range read (DS-MRR)	Yes		Yes
DS-MRR with Key-ordered retrieval	Yes		
Index_merge / Sort_intersection	Yes		
Cost-based choice of range vs. index_merge	Yes		
ORDER BY ... LIMIT <small_limit>	In 10.0		Yes
Use extended (hidden) primary keys for innodb/xtradb	5.5		
Batched key access (BKA)	Yes		Yes
Block hash join	Yes		
User-set memory limits on join buffers	Yes		
Apply early outer table ON conditions	Yes		
Null-rejecting conditions tested early for NULLs	Yes		

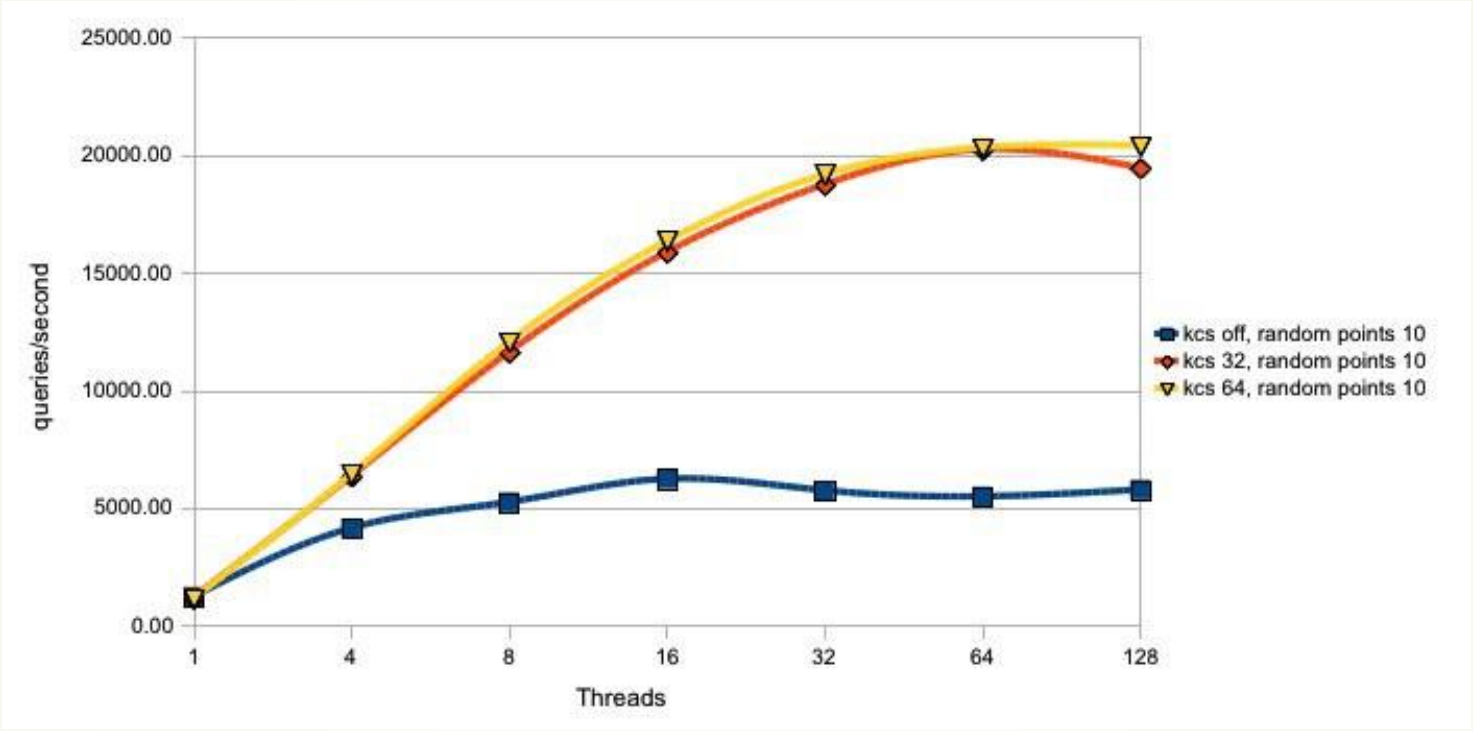
Optimizations comparison

Features	MariaDB 5.3/5.5	MySQL 5.5	MySQL 5.6
Subquery: In-to-exists	Yes	Yes	Yes
Subquery: Semi-join	Yes		Yes
Subquery: Materialization	Yes		Yes
Subquery: NULL-aware Materialization	Yes		
Subquery: Cost choice of materialization vs. in-to-exists	Yes		
Subquery: Cache	Yes		
Subquery: Fast explain with subqueries	Yes		

Optimizations comparison

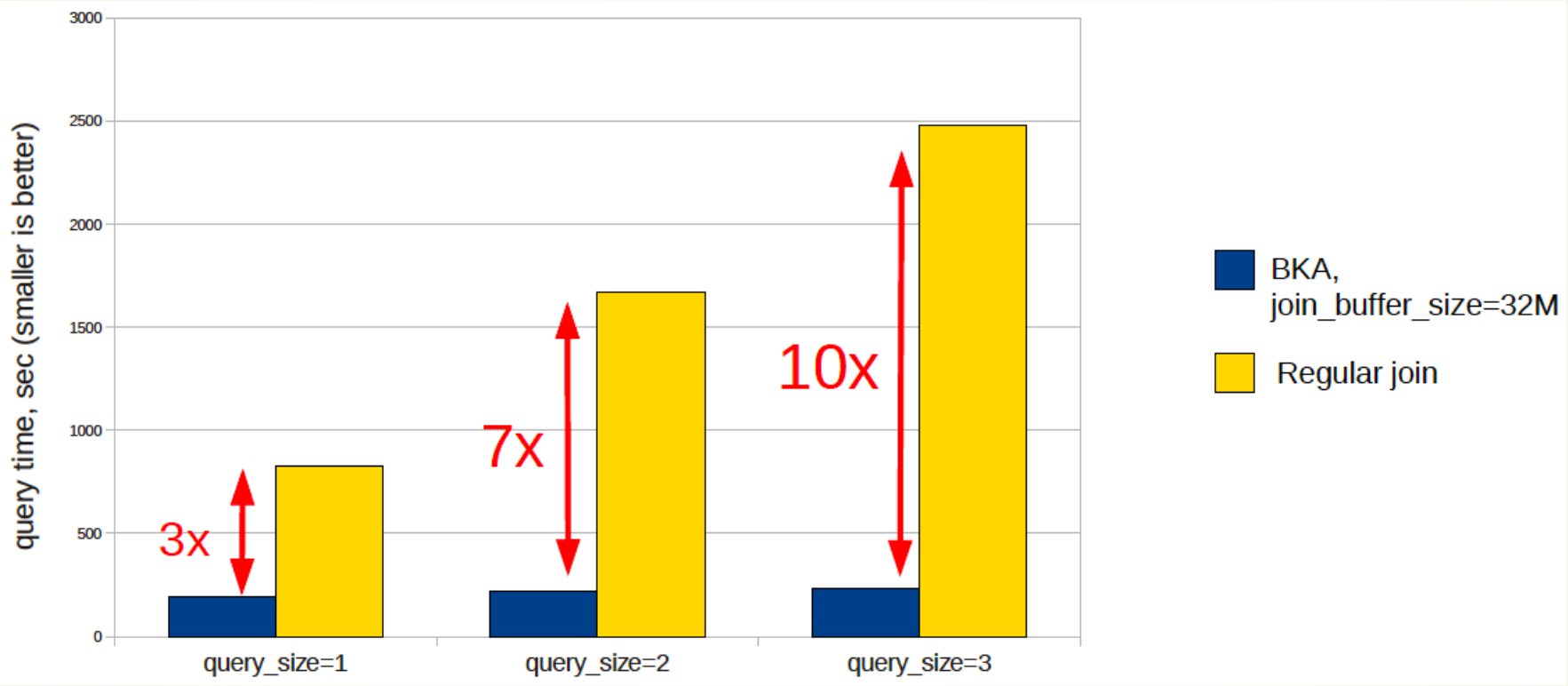
Features	MariaDB 5.3/5.5	MySQL 5.5	MySQL 5.6
Delayed materialization of derived tables / materialized views	Yes		Yes
Instant EXPLAIN for derived tables	Yes		Yes
Derived Table with Keys optimization	Yes		Yes
Fields of merge-able views and derived tables used in equality optimizations	Yes		Yes
LIMIT ROWS EXAMINED rows_limit	5.5		
Systematic control of all optimizer strategies	Yes		Partial
Explain for DELETE, INSERT, REPLACE, and UPDATE			Yes
EXPLAIN in JSON format			Yes
More detailed and consistent EXPLAIN for subqueries	Yes		

MyISAM Segmented key cache



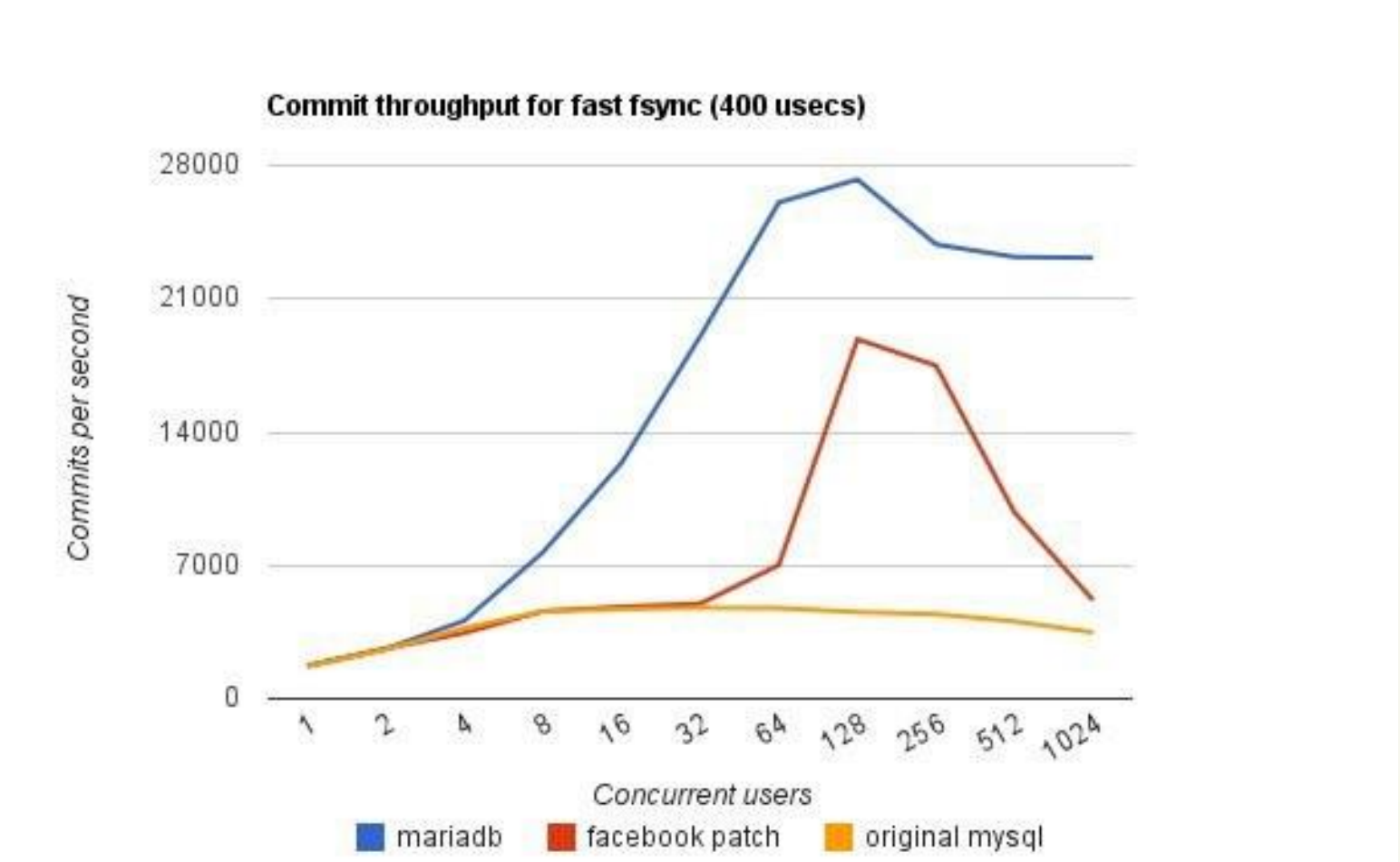
- Blue line is without segmented key cache.
- Solves one of the major read bottlenecks for MyISAM
- We see up to 250% performance gain depending on the amount of concurrent users.
- Fix applies to all MyISAM usage with many readers!

New Batched Key Access Speedups



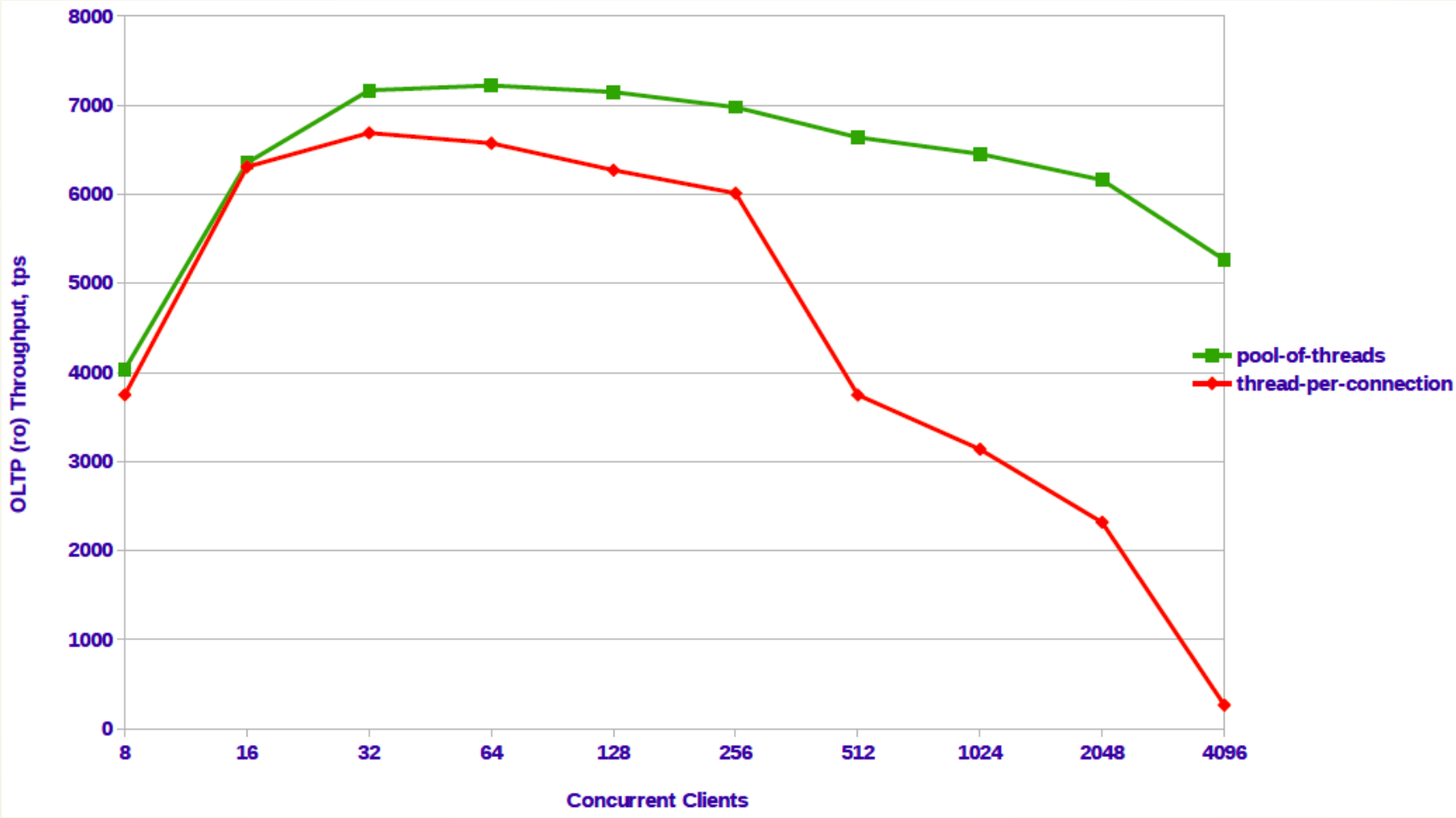

```
select max(l_extendedprice) from orders, lineitem where
o_orderdate between $DATE1 and $DATE2 and l_orderkey=o_orderkey
```

Group commit, verified



Source: Mark Callaghan's facebook blog for a server with 400 microsecond fsync latency

New thread pool for 5.5

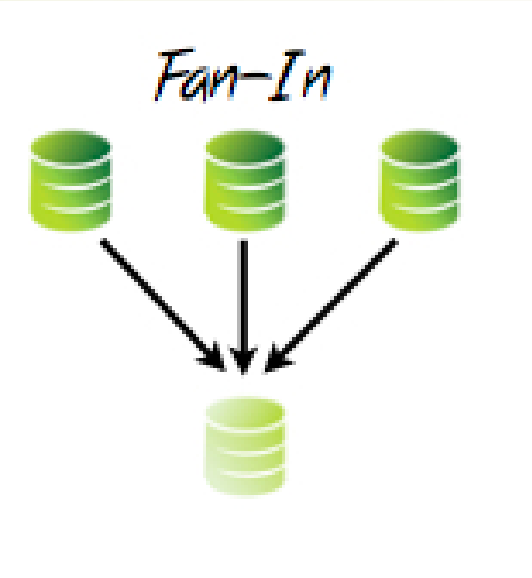


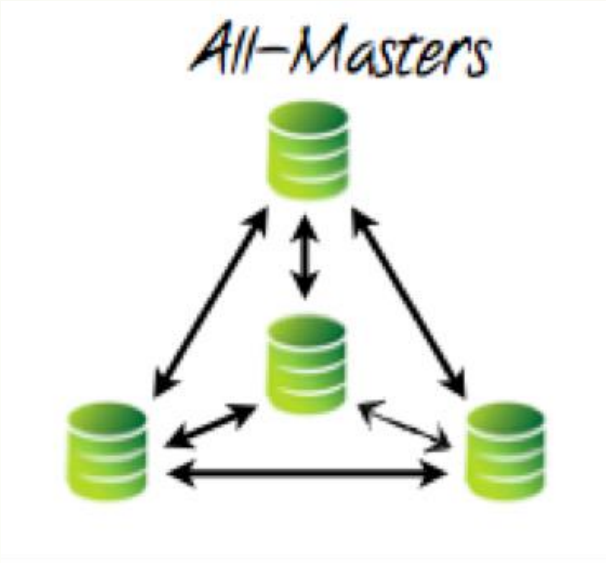
Why MariaDB 10.0

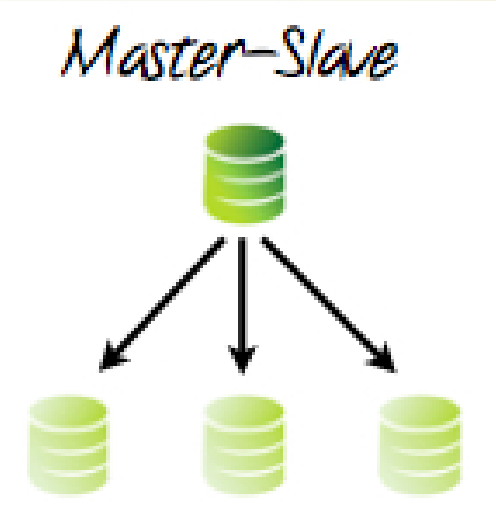
- MariaDB 5.5 already have most (+ a lot more) of the optimizer features of MySQL 5.6
- MariaDB 5.5 is already a superset of MySQL 5.5. MySQL 5.6 will only have a fraction of the MariaDB 5.5 new features.
- A full merge of MySQL 5.6 into MariaDB 5.6 is a one year project as a lot of the code has to be completely rewritten.
  - Features and usable code are removed, either intentionally or by mistake
  - New code is way to complex (you can do same thing in a fraction of the code)
  - It's clear that some of the new MySQL programmers doesn't understand the current code (see Kristian Nielsen's blog)
  - A lot of the new code is re-factoring we don't want to have.

  → Better to do the merge in 2 steps into 10.0, 10.1
- MariaDB 10.1 will have all important features of MySQL 5.6

Multi-Source Replication







- All Masters: Log_slave_updates = 0

Multi-Source Replication

```
root@localhost : (none) 03:30:26> show all slaves status\G
*************************** 1. row ***************************
               Connection_name: Emcon
              Slave_SQL_State: Slave has read all relay log; waiting
               Slave_IO_State: Waiting for master to send event
                  Master_Host: 10.0.3.29
                  Master_User: repl
                  Master_Port: 3306
                Connect_Retry: 60
              Master_Log_File: mysql-bin.001328
          Read_Master_Log_Pos: 530535455
               Relay_Log_File: mysql-relay-bin-emcon.001350
                Relay_Log_Pos: 530535740
        Relay_Master_Log_File: mysql-bin.001328
             Slave_IO_Running: Yes
*************************** 2. row ***************************
               Connection_name: owmn
              Slave_SQL_State: Slave has read all relay log; waiting
               Slave_IO_State: Waiting for master to send event
                  Master_Host: 10.0.3.21
                  Master_User: repl
                  Master_Port: 3306
                Connect_Retry: 60
              Master_Log_File: mysql-bin.000327
          Read_Master_Log_Pos: 14536305
               Relay_Log_File: mysql-relay-bin-owmn.000184
                Relay_Log_Pos: 14536590
        Relay_Master_Log_File: mysql-bin.000327
             Slave_IO_Running: Yes
            Slave_SQL_Running: Yes
```

Per-thread memory usage

```
| ID | USER            | HOST      | DB   | COMMAND | TIME | STATE                 | INFO                                                | MEM_USAGE |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
|  2 | root            | localhost | NULL | Query   |    0 | executing             | select * from information_schema.PROCESSLIST        |     55200 |
|  1 | event_scheduler | localhost | NULL | Daemon  |    6 | Waiting on empty queue | NULL                                                |     13544 |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
2 rows in set (0.00 sec)

root@localhost : (none) 11:01:33> select * from information_schema.PROCESSLIST;
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
| ID | USER            | HOST      | DB   | COMMAND | TIME | STATE                 | INFO                                                | MEM_USAGE |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
|  2 | root            | localhost | NULL | Query   |    0 | executing             | select * from information_schema.PROCESSLIST        |     55200 |
|  1 | event_scheduler | localhost | NULL | Daemon  |    6 | Waiting on empty queue | NULL                                                |     13544 |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
2 rows in set (0.00 sec)

root@localhost : (none) 11:01:33> select * from information_schema.PROCESSLIST;
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
| ID | USER            | HOST      | DB   | COMMAND | TIME | STATE                 | INFO                                                | MEM_USAGE |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
|  2 | root            | localhost | NULL | Query   |    0 | executing             | select * from information_schema.PROCESSLIST        |     55200 |
|  1 | event_scheduler | localhost | NULL | Daemon  |    7 | Waiting on empty queue | NULL                                                |     13544 |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
2 rows in set (0.00 sec)

root@localhost : (none) 11:01:34> select * from information_schema.PROCESSLIST;
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
| ID | USER            | HOST      | DB   | COMMAND | TIME | STATE                 | INFO                                                | MEM_USAGE |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
|  3 | root            | localhost | test | Query   |    3 | Sorting result        | select * from (select * from t1 a order by a.c)t order by c |   2303560 |
|  2 | root            | localhost | NULL | Query   |    0 | executing             | select * from information_schema.PROCESSLIST        |     55264 |
|  1 | event_scheduler | localhost | NULL | Daemon  |   21 | Waiting on empty queue | NULL                                                |     13544 |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
3 rows in set (0.00 sec)

root@localhost : (none) 11:01:48> select * from information_schema.PROCESSLIST;
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
| ID | USER            | HOST      | DB   | COMMAND | TIME | STATE                 | INFO                                                | MEM_USAGE |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
|  3 | root            | localhost | test | Query   |    8 | Sending data          | select * from (select * from t1 a order by a.c)t order by c |  62929760 |
|  2 | root            | localhost | NULL | Query   |    0 | executing             | select * from information_schema.PROCESSLIST        |     55264 |
|  1 | event_scheduler | localhost | NULL | Daemon  |   26 | Waiting on empty queue | NULL                                                |     13544 |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
3 rows in set (0.00 sec)

root@localhost : (none) 11:01:53> select * from information_schema.PROCESSLIST;
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
| ID | USER            | HOST      | DB   | COMMAND | TIME | STATE                 | INFO                                                | MEM_USAGE |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
|  3 | root            | localhost | test | Sleep   |    9 |                       | NULL                                                |     52184 |
|  2 | root            | localhost | NULL | Query   |    0 | executing             | select * from information_schema.PROCESSLIST        |     55216 |
|  1 | event_scheduler | localhost | NULL | Daemon  |   27 | Waiting on empty queue | NULL                                                |     13544 |
+----+-----------------+-----------+------+---------+------+-----------------------+-----------------------------------------------------+-----------+
3 rows in set (0.00 sec)

root@localhost : (none) 11:01:54>
```

There are a lot of others involved

- Most features in MariaDB 5.2 were contributed by the community!
- Many of the advanced features in MariaDB 5.3 are sponsored features
- In the askmonty.org knowledge base (free MariaDB and MySQL documentation) we have now 2800+ articles (mostly English)

Statistics from the past month:

- Added/Changed Articles: 201
- On Freenode #maria, 279 people wrote 6144 lines
- Launchpad Activity:
  - 27 active branches
  - 313 commits
- Hundreds of thousands of downloads of MariaDB. Probably >> 1M users
- We have seen companies converting hundreds of machines to MariaDB in a few days without any problems.
  - See http://kb.askmonty.org/en/mariadb-case-studies
- Distributions are moving from having included MariaDB to make

MariaDB popularity is increasing

- In December 2012
  - Wikipedia announced they are moving to MariaDB. Some parts has already moved.
- In January 2013
  - DB at Mozilla blogged they have moved to MariaDB
  - A Google developer said on #maria channel that Google is moving to MariaDB
  - Fedora voted 7-0 to make MariaDB the default MySQL database on Fedora.
  - OpenSuse announced that they have also made MariaDB default.
  - Chakra Linux announced that they have also made MariaDB default.

New LGPL client libraries

- LGPL client libraries for C and Java
  - C is based on the LGPL library from MySQL 3.23
  - API compatible with latest MySQL client libraries.
  - Java is based on the drizzle driver.
- Works with MariaDB, Percona server, MySQL and drizzle
- Developed by Monty Program Ab and SkySQL.
- Announced and released 2012-10-29

- You can download these from http://mariadb.org

- Documentation is still in progress...

MariaDB and TokuDB

MariaDB and Tokutek have agreed to make TokuDB a native plugin in MariaDB 5.5 and MariaDB 10.0 by end of Q1 2013.

This means that the official MariaDB binary will be able to dynamically load the TokuDB storage engine directly (no patches needed for MariaDB).

TokuDB will be added to the MariaDB buildbot test suite to ensure that the combination is properly tested on all supported platforms.
TokuDB will be available for download from https://downloads.mariadb.org/ together with MariaDB.

About TokuDB
- TokuDB uses Fractal Tree. indexing to improve insert and query speed, compression, replication performance, and online schema flexibility.
- TokuDB is created by Tokutek Inc. See www.tokutek.com for

Connect storage engine

MariaDB 10.0 will include the Connect storage engine by Olivier Betrand.

With the connect storage engine you can read, write and update files in a lot of different storage formats:

- Various fixed and dynamic text formats
- .DBF (dBASE format)
- .CSV
- .INI
- .XML
- ODBC ; Table extracted from an application accessable with ODBC

MariaDB Foundation Overview

The Foundation is the new driver of the MariaDB project

Custodian of the code, Guardian of the community

Foundation can never to be controlled by a single entity or person

Foundation designed to be self-sustaining

MariaDB Foundation Goals

That MariaDB be actively developed in the community and to:

- Increase adoption of MariaDB
- Ensure sustainable high-quality efforts to build, test and distribute MariaDB
- Ensure that community patches are reviewed and adopted
- Guarantee a community voice
- Keep MariaDB compatible with MySQL
- Maintain mariadb.org

MariaDB Foundation

More founders and sponsors are welcome!

If you care about the future of the MySQL ecosystem, please contact us and and ask how you can get involved!

Niall McCarthy mccarthy@emerge-open.com
Michael Widenius monty@mariadb.org
Andrew Katz andrew@mariadb.org

Conclusions

- MariaDB is maintained by the people that originally created MySQL and has the best knowledge of the MySQL code.
- MariaDB is binary compatible with MySQL, so its trivial to replace MySQL with MariaDB (minutes).
- Reasons to switch to MariaDB
  - Faster queries thanks to XtraDB (InnoDB plugin fork from Percona), a better optimizer and replication and better code.
  - Open source development: Anyone can be part of the development at all stages. Dev meetings are public.
  - More features, including critical ones like microseconds, multi-source and dynamic column support.
  - Less risk as MariaDB will not remove features like MySQL is doing (thread pool, storage engines, safemalloc (developer feature), etc)

Questions?

For developer questions later, use the public MariaDB email list at maria-discuss@lists.launchpad.net or IRC #maria on Freenode.

You can reach me for anything at monty@mariadb.org

欢迎莅临

2013中国数据库技术大会