MariaDB Server 10.2: The *somewhat* Complete Guide

Colin Charles, Chief Evangelist, Percona Inc
colin.charles@percona.com / byte@bytebot.net
http://bytebot.net/blog/ | @bytebot on Twitter
Originally at Percona Live Santa Clara, California, USA
on 26 April 2017, now as a webminar on 27 June 2017

License

- Creative Commons BY-NC-SA 4.0
- https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

whoami

- Chief Evangelist (in the CTO office), Percona Inc
- Founding team of MariaDB Server (2009-2016), previously at Monty Program Ab, merged with SkySQL Ab, now MariaDB Corporation
- Formerly MySQL AB (exit: Sun Microsystems)
- Past lives include Fedora Project (FESCO), OpenOffice.org
- MySQL Community Contributor of the Year Award winner 2014

What is MariaDB Server?

MariaDB Server is one of the most popular database servers in the world. It's made by the original developers of MySQL and guaranteed to stay open source. Notable users include Wikipedia, WordPress.com and Google.

MariaDB turns data into structured information in a wide array of applications, ranging from banking to websites. It is an enhanced, drop-in replacement for MySQL. MariaDB is used because it is fast, scalable and robust, with a rich ecosystem of storage engines, plugins and many other tools make it very versatile for a wide variety of use cases.

MariaDB is developed as open source software and as a relational database it provides an SQL interface for accessing data. The latest versions of MariaDB also include GIS and JSON features.

What is MariaDB Server, *really?*

- GPLv2 **fork** of MySQL 5.5 with lots of added **community** development
- Different features
- **Generally application compatible** with aims for feature completeness with MySQL
- A somewhat “drop-in replacement” (from an upgrade standpoint)

MariaDB thru the years

- MariaDB 5.1
- MariaDB 5.2
- MariaDB 5.3
- MariaDB 5.5
- MariaDB 10.0
- MariaDB 10.1
- MariaDB 10.2
- MariaDB 10.3 *alpha*
- MariaDB Galera Cluster 5.5
- MariaDB Galera Cluster 10.0
- C Connector
- Java Connector
- ODBC Connector

MariaDB governance

- MariaDB Corporation
  - Releases other products like MariaDB MaxScale, MariaDB ColumnStore
- MariaDB Foundation
  - Ensuring continuity and open collaboration in the MariaDB ecosystem

MariaDB Server 5.1

- Aria as the temporary table storage engine
- Thread pool v1
- Table Elimination - https://mariadb.com/kb/en/mariadb/what-is-table-elimination/
- Microsecond precision in processlist
  - `select id, time, time_ms, command, state from information_schema.processlist, (select sleep(2)) t;`

MariaDB Server 5.2

- Virtual columns (*ala* generated columns in MySQL 5.7)
- Pluggable authentication
- User statistics

MariaDB Server 5.3

- Optimiser, optimser, optimiser - https://mariadb.com/kb/en/mariadb/optimizer-feature-comparison-matrix/
- Many replication improvements that are only present in newer MySQL
- HandlerSocket
- Dynamic columns

MariaDB Server 5.5

- LIMIT ROWS EXAMINED
- non-blocking client API
- Threadpool v2
- MySQL 5.5 base

MariaDB Server 10.0

- replication: parallel replication, GTID, multi-source replication
- engines: cassandra, connect, spider, tokudb, mroonga
- show explain + output in slow query log
- roles, pcre regex

Why MariaDB?

- MySQL ecosystem development is at its most vibrant now than it has ever been
- Community can get features inside a shipping server with ease
- Storage engine vendors get shipping & wide distribution including testing
- Roadmaps are public on Jira

Where is MariaDB found?

- http://mariadb.org/
- Your Linux/BSD **distribution** - it is the default in RHEL 7, SUSE Enterprise 12, Debian 9, openSUSE, CentOS, Fedora, OpenBSD, etc.
- Pivotal.io CloudFoundry, RackSpace Cloud, Azure, Amazon AWS RDS, etc.

The 10.1 “community release”

- 30 Jun 2014 - MariaDB 10.1.0
- 17 Oct 2014 - MariaDB 10.1.1
- 7 Dec 2014 - MariaDB 10.1.2
- 2 March 2015 - MariaDB 10.1.3
- 13 April 2015 - MariaDB 10.1.4
- 4 June 2015 - MariaDB 10.1.5
- 27 July 2015 - MariaDB 10.1.6
- 9 September 2015 - MariaDB 10.1.7 RC
- 17 October 2015 - MariaDB 10.1.8 GA
- 23 Nov 2015, 24 Dec 2015, 29 Jan 2016, 25 Feb 2016, 25 March 2016, 10 May 2016, 18 July 2016, 30 Aug 2016, 30 Sep 2016, 7 Nov 2016, 15 Dec 2016, 18 Jan 2017, 14 Mar 2017

The 10.2 release so far

- 18 April 2016 - 10.2.0 alpha
- 4 July 2016 - 10.2.1 alpha
- 27 September 2016 - 10.2.2 beta
- 24 December 2016 - 10.2.3 beta
- 17 February 2017 - 10.2.4 release candidate
- 5 April 2017 - 10.2.5 release candidate
- 23 May 2017 - 10.2.6 stable GA

10.2 Themes

- Analytical queries, optimiser
- Client & protocol
- Removing historical limitations, refactoring
- Replication
- Engines: InnoDB, MyRocks

Analytical queries

- views and subqueries in the FROM clause
- Common Table Expressions + recursive CTEs
- Window functions
- Pushdown conditions into non-mergeable views/ derived tables

The analytical queries slide deck to visit

- Common Table Expressions and Window Functions simple, maintainable, fast queries by Vicentiu-Marian Ciorbaru

- 26 April - 2:00 PM - 2:50 PM @ Ballroom E

- https://www.percona.com/live/17/sessions/common-table-expressions-and-window-functions-simple-maintainable-fast-queries

Client & protocol

- MariaDB Connector/C for MySQL and MariaDB Server. It is libmysqlclient API compatible. LGPL. OpenSSL/GnuTLS/schannel (no more yaSSL/ wolfSSL)

- EXECUTE IMMEDIATE *statement*

Removing historical limitations, refactoring

- CHECK CONSTRAINT enforced (previously accepted but ignored)
  - https://mariadb.com/kb/en/mariadb/constraint/
- DEFAULT clause can be an arbitrary expression
  - https://mariadb.com/kb/en/mariadb/create-table/#default
- BLOB & TEXT columns can now have a DEFAULT value
- Generated columns - virtual columns can finally be indexed (not just persistent), up to 64K per expression. Constant expressions, refer to virtual columns, and it works with non-deterministic functions
- Supported decimals in DECIMAL is now 38 (was 30)

Replication

- Default binlog format is now MIXED
- Default replicate_annotate_row_events is ON
- Binlog event compression
- Time delayed replication arrives
- read_binlog_speed_limit - restricting the speed at which the slave reads the binlog from the master
- DML only Flashback - rollback instances/databases/tables to an older snapshot (via Alibaba!)
- Continuous streaming binary log backup added to mysqlbinlog

The Flashback slide deck to review

- Slides: https://www.percona.com/live/17/sessions/flashback-rolling-back-mysqlmariadb-instance-database-or-table-previous-snapshot

JSON & GeoJSON

- SQL Standard 2014, with MySQL compatible extensions and functions
- No JSON type - not in the standard (it says to use regular VARCHAR or TEXT); apparently there are benchmarks, and VARCHAR is as fast as JSON type (waiting to be published)

InnoDB from MySQL 5.7

- This is the first release of MariaDB Server that does not have Percona XtraDB as the default InnoDB!
- InnoDB is taken from MySQL 5.7
- New variable for setting a directory for storing temporary non-tablespace InnoDB files, innodb_tmpdir
- Persistent AUTO_INCREMENT for InnoDB

MyRocks

- This is integrated and a big deal!
- Alpha-maturity plugin at the moment

The MyRocks slide deck to review

- MariaRocks: MyRocks in MariaDB by Sergei Petrunia
- 27 April - 11:00 AM - 11:50 AM @ Ballroom E
- https://www.percona.com/live/17/sessions/mariarocks-myrocks-mariadb

Manageability

- CREATE USER MAX_*_PER_HOUR & MAX_USER_CONNECTIONS
- Limiting resource usage, and also TLS/SSL

Google Summer of Code

- SQL Roles
- Kerberos authentication
- PCRE regular expressions
- (per-query variable settings)

10.1 Themes

- Security
- High Availability
- High Performance
- Operational Ease
- Better for developers and DBAs

Security

Encryption

- Encryption: tablespace and table level encryption with support for rolling keys using the AES algorithm
  - table encryption — `PAGE_ENCRYPTION=1`
  - tablespace encryption — encrypts everything including log files
- New `file_key_management_filename`, `file_key_management_filekey`, `file_key_management_encryption_algorithm`
- Well documented — https://mariadb.com/kb/en/mariadb/data-at-rest-encryption/

Encryption II

- The key file contains encryption keys identifiers (32-bit numbers) and hex-encoded encryption keys (128-256 bit keys), separated by a semicolon.
- don’t forget to create keys!
  - eg. `openssl enc -aes-256-cbc -md sha1 -k secret -in keys.txt -out keys.enc`

Encryption III

```
CREATE TABLE customer (
   customer_id bigint not null primary key,
   customer_name varchar(80),
   customer_creditcard varchar(20))
ENGINE=InnoDB

page_encryption=1

page_encryption_key=1;
```

Encryption IV

- Tablespace encryption (Google)
  - again, you need to pick an encryption algorithm
  - specify what to encrypt: `innodb-encrypt-tables, aria, aria-encrypt-tables, encrypt-tmp-disk-tables, innodb-encrypt-log`
- don’t forget key rotation:
  - `innodb-encryption-threads=4`
  - `innodb-encryption-rotate-key-age=1800`

Encryption V

- `/etc/my.cnf.d/enable_encryption.preset`
- Consider using Eperi Gateway for Databases
- MariaDB Enterprise has a plugin for Amazon Key Management Server (KMS)
  - You can also *compile* this yourself as the source is provided
- mysqlbinlog has no way to read (i.e. decrypt) an encrypted binlog
- This does not work with MariaDB Galera Cluster yet (gcache is not encrypted yet), and also xtrabackup needs additional work (i.e. if you encrypt the redo log)

Password validation

- simple_password_check password validation plugin
  - can enforce a minimum password length and guarantee that a password contains at least a specified number of uppercase and lowercase letters, digits, and punctuation characters.
- cracklib_password_check password validation plugin
  - Allows passwords that are strong enough to pass CrackLib test. This is the same test that pam_cracklib.so does

SQL Error Logging Plugin

- Log errors sent to clients in a log file that can be analysed later. Log file can be rotated (recommended)
- a MYSQL_AUDIT_PLUGIN

```
install plugin SQL_ERROR_LOG soname
'sql_errlog.so';
```

Audit Plugin

- Log server activity - who connects to the server, what queries run, what tables touched - rotating log file or syslogd
- MariaDB has extended the audit API, so user filtering is possible
- a MYSQL_AUDIT_PLUGIN

```
INSTALL PLUGIN server_audit SONAME
‘server_audit.so’;
```

Authentication Plugins

- Shipped a PAM authentication plugin for the longest time
- Now you also have a Kerberos/GSSAPI based authentication plugin
  - works with Microsoft Active Directory also
- ed25519 alternative to mysql_native_password

High Availability

MariaDB 10 replication: provisioning a new slave

```
SET GLOBAL GTID_SLAVE_POS =
BINLOG_GTID_POS("masterbin.00045",
600);

CHANGE MASTER TO
master_host="192.168.2.4",
master_use_gtid=slave_pos;

START SLAVE;
```

Multi-source replication

- Work from Taobao/Alibaba
- Many users partition data across many masters... now you can replicate many masters to a single slave
- Great for analytical queries, complete backups, etc.
- All master/slave commands take a connection name now (like CHANGE MASTER “connection_name”, SHOW SLAVE “connection_name” STATUS, etc.)

Galera Cluster integrated

- Full integration of Galera Cluster into MariaDB 10.1 — not a separate download

- Granular monitoring in INFORMATION_SCHEMA — WSREP_MEMBERSHIP, WSREP_STATUS

Optimistic parallel replication

- Before, transactions committed in parallel on the master could be run in parallel
- Now, more than one transaction will be considered to be run in parallel giving another performance boost in master-to-slave replication
- Need a 10.1 master to work
- https://mariadb.atlassian.net/browse/MDEV-6676

The parallel replication slides to review

- MySQL/MariaDB Parallel Replication: inventory, use cases and limitations by Jean-François Gagné

- 26 April - 1:00 PM - 1:50 PM @ Ballroom E

- https://www.percona.com/live/17/sessions/mysqlmariadb-parallel-replication-inventory-use-cases-and-limitations

Replication: START TRANSACTION WITH CONSISTENT SNAPSHOT

- Works with the binlog, possible to obtain the binlog position corresponding to a transactional snapshot of the database without blocking any other queries.
    - by-product of group commit in the binlog to view commit ordering
- Used by the command `mysqldump--single-transaction --master-data` to do a fully non-blocking backup which can be used to provision a new slave
- Works consistently between transactions involving more than one storage engine
- https://kb.askmonty.org/en/enhancements-for-start-transaction-with-consistent/

More in replication

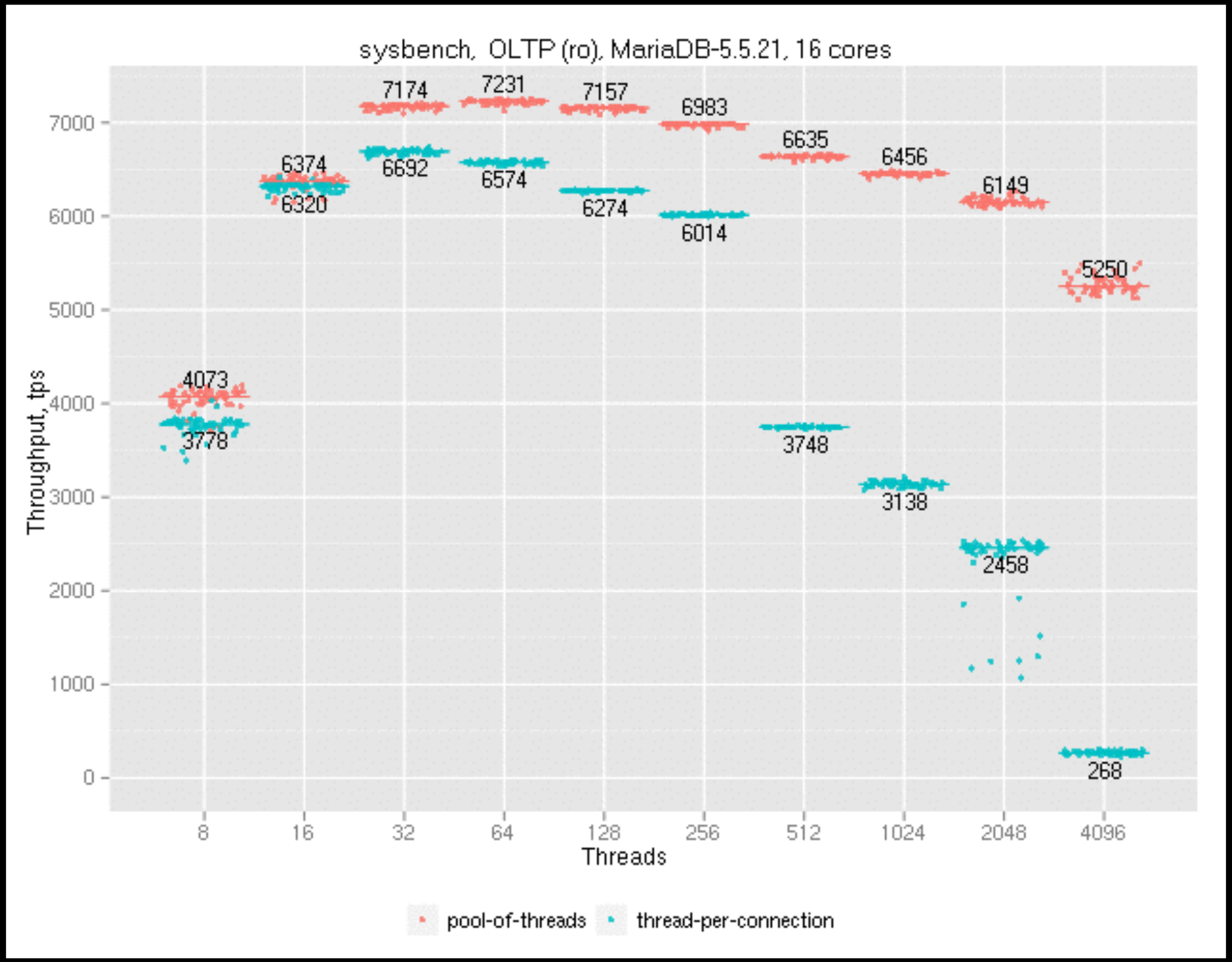
- Enhanced semi-sync replication (like in 5.7 - thanks FB/Google)
- `domain_id` based replication filters
- Slaves can execute triggers now (in RBR)
- Dump thread enhancements (remove binlog lock LOCK_log) from 5.7 included (Google)

High Performance

An opensource threadpool

- Modified from 5.1 (libevent based), great for CPU bound loads and short running queries
- No minimization of concurrent transactions with dynamic pool size
- `thread_handling=pool-of-threads`
- https://mariadb.com/kb/en/mariadb/thread-pool-in-mariadb/
- now you can also have a priority mode for tickets

Threadpool

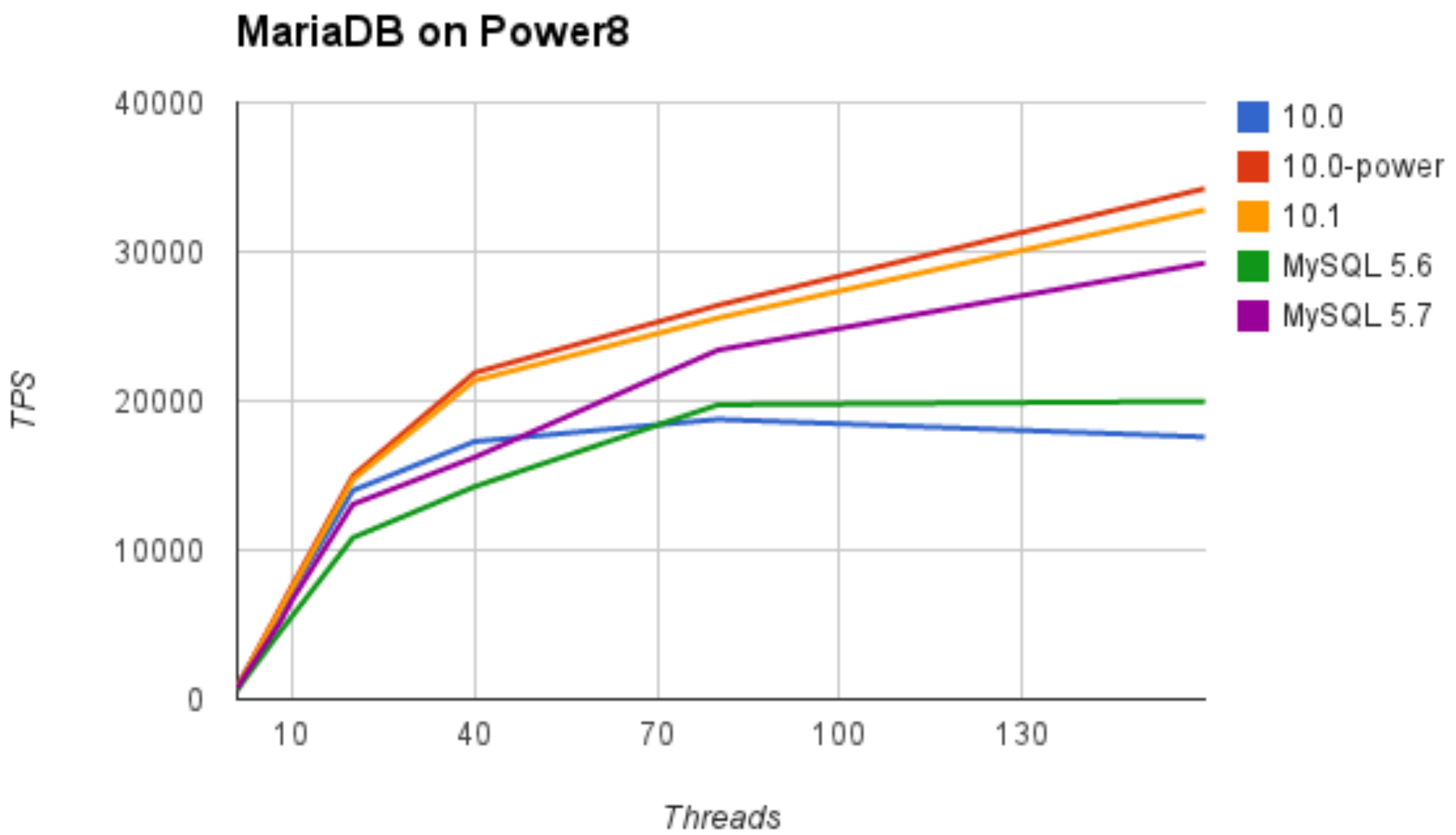

sysbench, OLTP (ro), MariaDB-5.5.21, 16 cores
Throughput, tps
Threads
7000
6000
5000
4000
3000
2000
1000
0
8
16
32
64
128
256
512
1024
2048
4096
7174
7231
7157
6983
6635
6456
6149
5250
6374
6320
6692
6574
6274
6014
4073
3778
3748
3138
2458
268
pool-of-threads
thread-per-connection

InnoDB improvements

- Multi-threaded flush (also in 5.7, different implementation)
- 64KB pages in InnoDB (old limit = 16KB).
- Defragementation (FB, backported by DaumKakao)
- `I_S.INNODB_SEMAPHORE_WAITS, I_S.INNODB_MUTEXES`
- Forced primary key
  - If option is true, create table without primary key or unique key where all keyparts are NOT NULL is not accepted. Instead an error message is printed.

InnoDB WebScaleSQL

- MDEV-6936: Buffer pool list scan optimisation
- MDEV-6929: Port Facebook Prefix Index Queries Optimization
- MDEV-6932: Enable Lazy Flushing
- MDEV-6931: Page cleaner should do LRU flushing regardless of server activity
- fixes mysql#71988, mysql#70500
- DB-746 merge clustering key is covering key for mariadb 10 (TokuDB)
- MDEV-6933: Spurious lock_wait_timeout_thread wakeup in lock_wait_suspend_thread()
  - fixes mysql#72123



- http://svoj-db.blogspot.ru/2014/12/mariadb-on-power8-2014-wrap-up.html

Operational Ease

Per query variables

- Long history (http://www.bytebot.net/blog/archives/2014/05/04/per-query-variable-settings-in-mysqlpercona-serverwebscalesql)
- `SET STATEMENT max_statement_time=1000 FOR SELECT name FROM name ORDER BY name;`

Statement timeouts

- `MAX_STATEMENT_TIME` to abort long running queries
  - Called “query timeouts” + have a compatible syntax
  - https://mariadb.atlassian.net/browse/MDEV-4427

Enhanced KILL syntax

- HARD | SOFT & USER USERNAME are MariaDB-specific (5.3.2)
- KILL QUERY ID query_id (10.0.5) - kill by query id, rather than thread id
- SOFT ensures things that may leave a table in an inconsistent state aren't interrupted (like REPAIR or INDEX creation for MyISAM or Aria)

```
KILL [HARD | SOFT] [CONNECTION | QUERY]
[thread_id | USER user_name]
```

Progress reporting

- ALTER TABLE & LOAD DATA INFILE

```
MariaDB [mail]> alter table mail engine = maria;

Stage: 1 of 2 'copy to tmp table' 17.55% of stage done

MariaDB [mail]> select id, user, db, command, state,
-> time_ms, progress from information_schema.processlist;

+--------+-------------------+----------+----------+
| command | state             | time_ms   | progress |
+--------+-------------------+----------+----------+
| Query   | copy to tmp table | 23407.131 |   17.551 |
+--------+-------------------+----------+----------+
1 row in set (0.47 sec)
```

INFORMATION_SCHEMA.PROCESSLIST

```
MariaDB [(none)]> select * from INFORMATION_SCHEMA.processlist, (select sleep(2)) t\
G
*************************** 1. row ***************************
           ID: 14
         USER: root
         HOST: localhost
           DB: NULL
      COMMAND: Query
         TIME: 2
        STATE: Filling schema table
         INFO: select * from INFORMATION_SCHEMA.processlist, (select sleep(2)) t
      TIME_MS: 2000.818
        STAGE: 0
    MAX_STAGE: 0
     PROGRESS: 0.000
  MEMORY_USED: 235984
EXAMINED_ROWS: 0
     QUERY_ID: 55
  INFO_BINARY: select * from INFORMATION_SCHEMA.processlist, (select sleep(2)) t
          TID: 21373
     sleep(2): 0
1 row in set (2.00 sec)
```

```
mysql [localhost] {msandbox} ((none)) > select * from INFORMATION_SCHEMA.process
list, (select sleep(2)) t\G

*************************** 1. row ***************************
      ID: 5
    USER: msandbox
    HOST: localhost
      DB: NULL
 COMMAND: Query
    TIME: 2
   STATE: executing
    INFO: select * from INFORMATION_SCHEMA.processlist, (select sleep(2)) t
sleep(2): 0
1 row in set (2.00 sec)
```

Better for Developers and DBAs

Optimiser enhancements

- UNION ALL without temporary tables (5.7)
- Improve ORDER BY … LIMIT in optimiser
- Histograms
- `ANALYZE <statement>`
- EXPLAIN JSON (like 5.6)*
- `EXPLAIN ANALYZE with FORMAT=JSON`
  - includes data from the query execution itself — this is MariaDB only
  - https://mariadb.com/kb/en/mariadb/analyze-formatjson-examples/

The Histogram slide deck to review

- Histograms in MySQL and MariaDB by Sergei Petrunia

- 26 April - 3:30 PM - 4:20 PM @ Ballroom E

- https://www.percona.com/live/17/sessions/histograms-mysql-and-mariadb

CONNECT

- CONNECT having full JSON/BSON support
  - Can read/write filename.json files with ease
- Writing — INSERT, UPDATE, DELETE is supported
- Naturally the other good use? ODBC connections to other databases

Other bits

- CREATE or REPLACE for most database objects minus indexes
- SET DEFAULT ROLE (there is a default role now for current user)
- FRM files are now not created for temporary tables
- `INFORMATION_SCHEMA.SYSTEM_VARIABLES` - information for system variables
- Microseconds in GET_LOCK()

Other bits

- Compiled with security hardening options (fortify source - https://mariadb.atlassian.net/browse/MDEV-5730)
- `@@sql_log_slow` can now be controlled on a session basis (not just globally)
- Sequence engine enabled by default
- `default_tmp_storage_engine` option
- ALGORITHM column in I_S.VIEWS

GIS

- Full compliance for the OGC standards around GIS.
- I_S tables: `GEOMETRY_COLUMN` `SPATIAL_REF_SYS`
- `REF_SYSTEM_ID per GEOMETRY` column
- See: MDEV-4045, MDEV-12, MDEV-60

Compatibility

- Temporary tables are stored in Aria but now there is a —`default-tmp-storage-engine` option
- `engine_condition_pushdown` flag removed (its always on for engines that support it)
- `--mysql56-temporal-format` option to use the MySQL-5.6 low level formats to store TIME, DATETIME and TIMESTAMP types
- PERFORMANCE_SCHEMA disabled by default like in 10.0
  - also, no .frm files created for faster startup
- see: https://mariadb.com/kb/en/mariadb/mariadb-vs-mysql-compatibility/

The 10.1 “Community Release”

- Google - encryption, scrubbing, enhanced semisync, dump thread, thd_specifics plugin service
- Eperi - encryption
- DaumKakao - defragmentation, online alter progress monitoring
- Antony Curtis - compound statements
- Sriram Patil (GSoC) - CREATE or REPLACE/IF NOT EXISTS
- Daniel Black - finer grained status variables for replication monitoring
- FusionIO - atomic writes, page compression, TRIM, multi-threaded flushing
- Facebook — defragmentation, prefix index queries optimization, lazy flushing, buffer pool list scan optimization, configurable long, semaphore wait timeout
- Percona - SET STATEMENT, enforce_storage_engine

Welcoming 5.7 features MariaDB had

- Multi-source replication
- Dynamic replication filters
- SHOW EXPLAIN for connection_id
- GIS functionality
- Statement timeouts
- Change master without stopping SQL thread
- Online GTID implementation
- GTID no longer requires `log-slave-updates` to be enabled
- Virtual columns (generated columns)
- SHUTDOWN command
- FusionIO functionality

MariaDB still *unique*! (vs. 5.7)

- Integrated Galera Cluster
- Table/tablespace encryption
- Optimistic parallel replication
- InnoDB defragmentation
- ANALYZE <statement>
- Threadpool
- cracklib_password_check
- SQL error logging plugin
- Extended REGEXP
- Roles

Participate!

- Contribute code - github.com/mariadb/server
- Write KB articles - http://mariadb.com/kb/
- Report bugs: http://mariadb.org/jira
- Join the merry gang on #maria at irc.freenode.net
- Enable the feedback plugin ([enable-feedback] in my.cnf)
- Mailing lists: {maria-discuss, maria-developers} @lists.launchpad.net
- Tweet at @mariadbfdn, + on GPlus

Books!

1. MariaDB Crash Course, Ben Forta (September 2011)
2. Getting Started with MariaDB, Daniel Bartholomew (October 2013)
3. MariaDB Cookbook, Daniel Bartholomew (March 2014)
4. Real MariaDB, Matt Lee (April 2014)
5. Building a Web Application with PHP & MariaDB: A Reference Guide, Sai Srinivas Sriparasa (June 2014)
6. MariaDB: Beginners Guide, Rodrigo Ribeiro (August 2014)
7. Mastering MariaDB, Federico Razzioli (September 2014)
8. MariaDB High Performance, Pierre Mavro (September 2014)
9. Learning MySQL & MariaDB, Russell Dyer (April 2015)

MariaDB Server 10.3

- Alpha release - 16 April 2017
- sql_mode=ORACLE seems to be the focus
- https://mariadb.com/kb/en/mariadb/what-is-mariadb-103/
- Maybe PROXY protocol support like Percona Server

In conclusion

- MariaDB is GPLv2 licensed
- Aims to be feature complete with MySQL + loaded with extras
    - Review compatibility docs: https://mariadb.com/kb/en/mariadb/mariadb-vs-mysql-compatibility/
- Enterprise features made open
- Wide distribution

Percona Live Europe Call for Papers & Registration are Open!

Championing Open Source Databases

- MySQL, MongoDB, Open Source Databases
- Time Series Databases, PostgreSQL, RocksDB
- Developers, Business/Case Studies, Operations
- September 25-27th, 2017
- Radisson Blu Royal Hotel, Dublin, Ireland

Submit Your Proposal by July 17th!

www.percona.com/live/e17

Thank you!

Colin Charles
colin.charles@percona.com / byte@bytebot.net
http://bytebot.net/blog | @bytebot on twitter
slides: slideshare.net/bytebot