Meet MariaDB: The new MySQL

Colin Charles, Team MariaDB, MariaDB Corporation
colin@mariadb.org / byte@bytebot.net
http://bytebot.net/blog/ | @bytebot on Twitter
http://mariadb.org/ | http://mariadb.com/
World Internet Developer Summit, Hong Kong
21 May 2015

whoami

- Work on MariaDB at MariaDB Corporation (~~SkySQL Ab~~)
  - Merged with Monty Program Ab, makers of MariaDB
- Formerly MySQL AB (exit: Sun Microsystems)
- Past lives include Fedora Project (FESCO), OpenOffice.org
- MySQL Community Contributor of the Year Award winner 2014

Global Top 20 Sites

1. Google
2. Facebook
3. YouTube
4. Baidu
5. Yahoo!
6. Wikipedia
7. Amazon
8. Twitter
9. QQ
10. Taobao
11. Google.co.in
12. LinkedIn
13. live.com
14. Sina
15. Weibo
16. yahoo.co.jp
17. Tmall
18. google.co.jp
19. google.de
20. Ebay

What is MariaDB Server?

- GPLv2 branch of MySQL with a lot of added **community** development
- Feature enhanced
- Application compatible & feature complete with MySQL
- A drop-in replacement


HAPPY
BIRTHDAY

MariaDB thru the years

- MariaDB 5.1
- MariaDB 5.2
- MariaDB 5.3
- MariaDB 5.5
- MariaDB 10.0
- MariaDB Galera Cluster 5.5
- MariaDB Galera Cluster 10.0
- C Connector
- Java Connector
- ODBC Connector

Why MariaDB?

- MySQL has a single owner; MariaDB has the MariaDB Foundation (not just Corporation)
- 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

Who's behind it?

- Many original MySQL developers including Michael "Monty" Widenius (MySQL's creator)
- Many contributors (with impressive community:corporation:foundation ratio)

MariaDB, the ecosystem

- Besides the Server, we do develop LGPL Connectors
- Focus on making “enterprise” features opensource
  - threadpool is a great example of working in the open
- Open content (& friendly licensed) Knowledge Base

MariaDB MaxScale

- Level 7 proxy router, that understands the MySQL protocol, with a pluggable architecture
- Possibilities are endless - use it for logging, writing to other databases (besides MySQL), preventing SQL injections via regex filtering, route via hints, query rewriting, have a binlog relay, etc.
- Load balance your Galera clusters

Where is MariaDB found?

- http://mariadb.org/
- Your Linux distribution - it is the default in RHEL 7, SUSE Enterprise 12, openSUSE, CentOS, Fedora, etc.
  - and a choice in Debian & Ubuntu

Using MariaDB?

Google

- A migration to MariaDB 10
- Funded parallel replication
- Code reviews (on public mailing list), QA, testing, etc.
- Contributed tablespace encryption in MariaDB 10.1

Wikipedia

- Wikipedia went from MySQL 5.1, to MariaDB 5.5, and now are already on MariaDB 10.0

- Initially, reasoning was optimizer enhancements, XtraDB, Foundation backed, bugs all fixed responsively

- Wikipedia improved query times: 4-15% faster

KakaoTalk

- Sped up messaging over 5x
- 48m monthly active users, with over 130m in DB, powered by MariaDB 10.0
- Code contributions — InnoDB defragmentation, backports from WebScaleSQL

Innovation today

- Multi-source replication
- GIS functionality
- NoSQL APIs (HandlerSocket), functionality via dynamic columns
- Parallel replication
- GTID done better
- CONNECT engine to read JSON/BSON

Source: Gartner (October 2014)

Galera Cluster integrated into 10.1

- Full integration of Galera Cluster into MariaDB 10.1 — it won't be a separate download
  - no lost transactions, optimisations for WAN replication, non-blocking DDL, no limits on transaction size
- `Server version: 10.1.3-MariaDB-wsrep MariaDB Server, wsrep_25.10.r4144`
- Granular monitoring in INFORMATION_SCHEMA — WSREP_MEMBERSHIP, WSREP_STATUS

Encryption

- Encryption: tablespace and table level encryption with support for rolling keys using the AES algorithm
    - tablespace encryption — encrypts everything including log files (not the binlog — yet)
        - tablespace scrubbing: background process that regularly scans through the tables and upgrades the encryption keys
- Overhead of ~10%
- XtraDB/InnoDB only; Aria for temporary tables
- New plugin: `file_key_management`

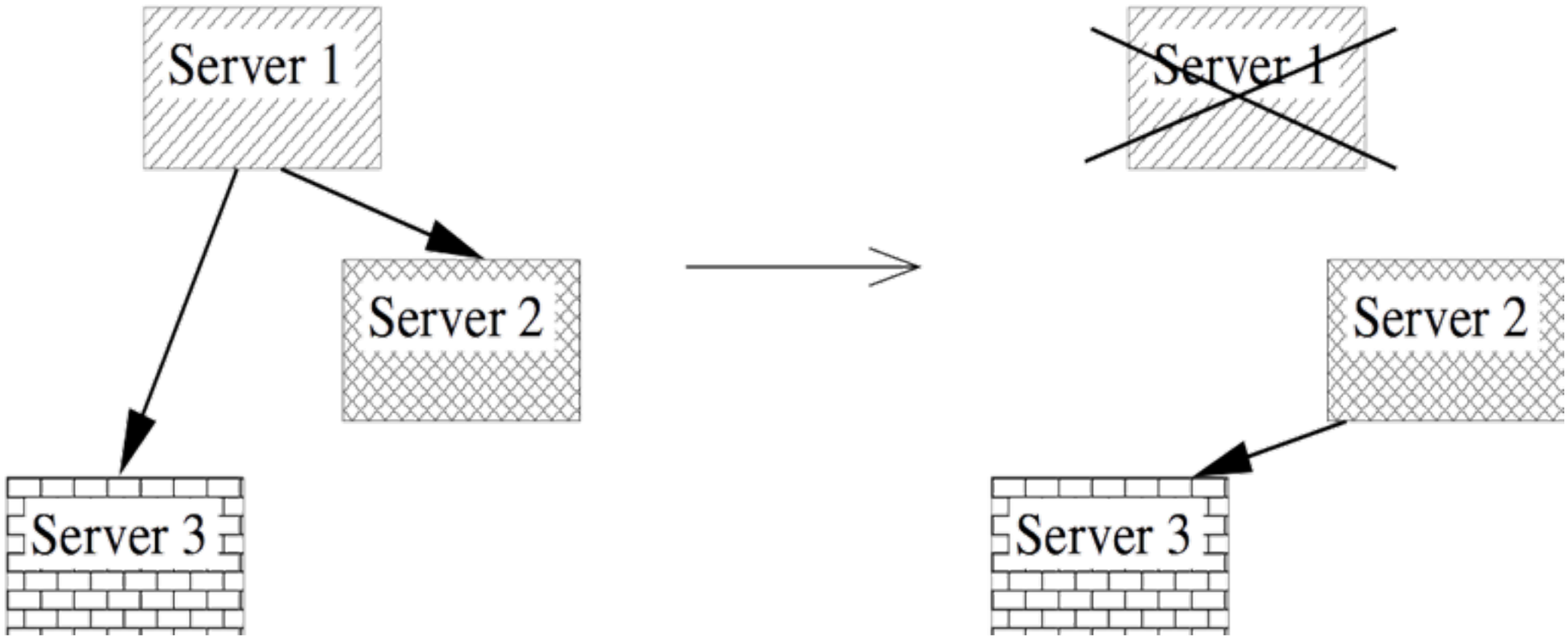
Encryption II

```
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;
```

MariaDB 10 replication

- Global Transaction ID
  - have complex replication topologies; simple failover & slave promotion
  - doesn't require restarts!
    - new slave provisioning: 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;
    - turning on GTID for slaves: STOP SLAVE CHANGE MASTER TO master_use_gtid=current_pos; START SLAVE;
    - change masters: STOP SLAVE CHANGE MASTER TO master_host="10.2.3.5"; START SLAVE;
- Crash-safe slaves - GTID position stored in InnoDB table

Automatic binlog position for master failover



- On Server2: CHANGE MASTER TO master_host='server2', master_use_gtid=1;

Why different GTID compared to 5.6?

- MySQL 5.6 GTID does not support multi-source replication
- Supports —log-slave-updates=0 for efficiency
- Enabled by default, with self-healing capabilities

Binlog (size matters!)

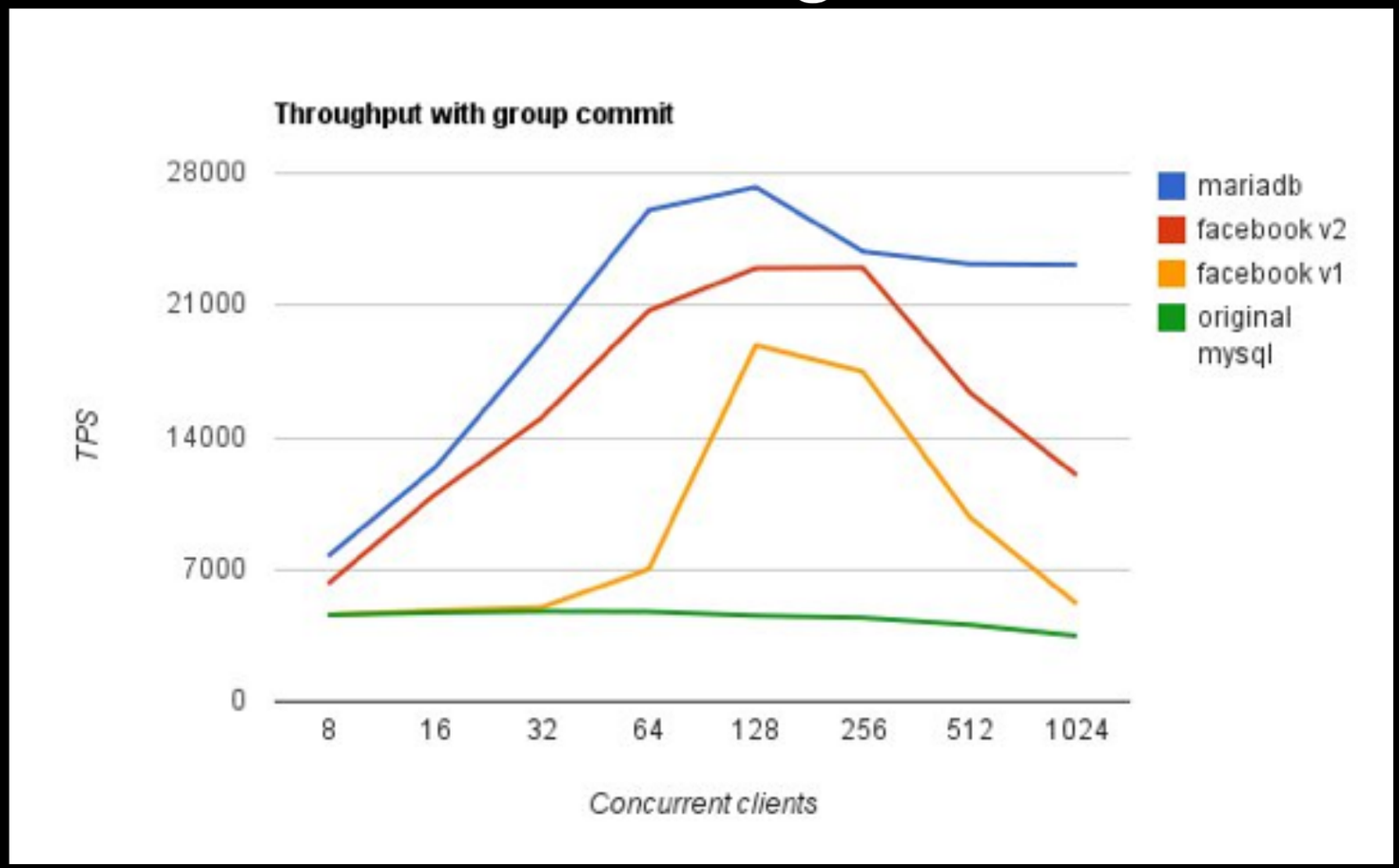
- Example query: INSERT INTO t1 VALUES (10, “foo”);
- MySQL 5.6… 265 bytes
- MariaDB 10.0… 161 bytes
- Do you want a 60% larger binlog size?

Multi-source replication

- Work from Taobao
- Many users partition data across many masters... now you can replicate many masters to a single slave
- Great for analytical queries, complete backups, etc.
- @@default_master_connection contains current connection name (used if connection name is not given)
- All master/slave commands take a connection name now (like CHANGE MASTER “connection_name”, SHOW SLAVE “connection_name” STATUS, etc.)
- https://kb.askmonty.org/en/multi-source-replication/

Replication: group commit in the binary log

- `sync_binlog=1, innodb_flush_log_at_trx_commit=1`
- SHOW STATUS LIKE 'binlog_%commits';



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
- We have to check if this only will work with a 10.1 master
- Isn’t fully pushed (or documented) yet — see: https://mariadb.atlassian.net/browse/MDEV-6676

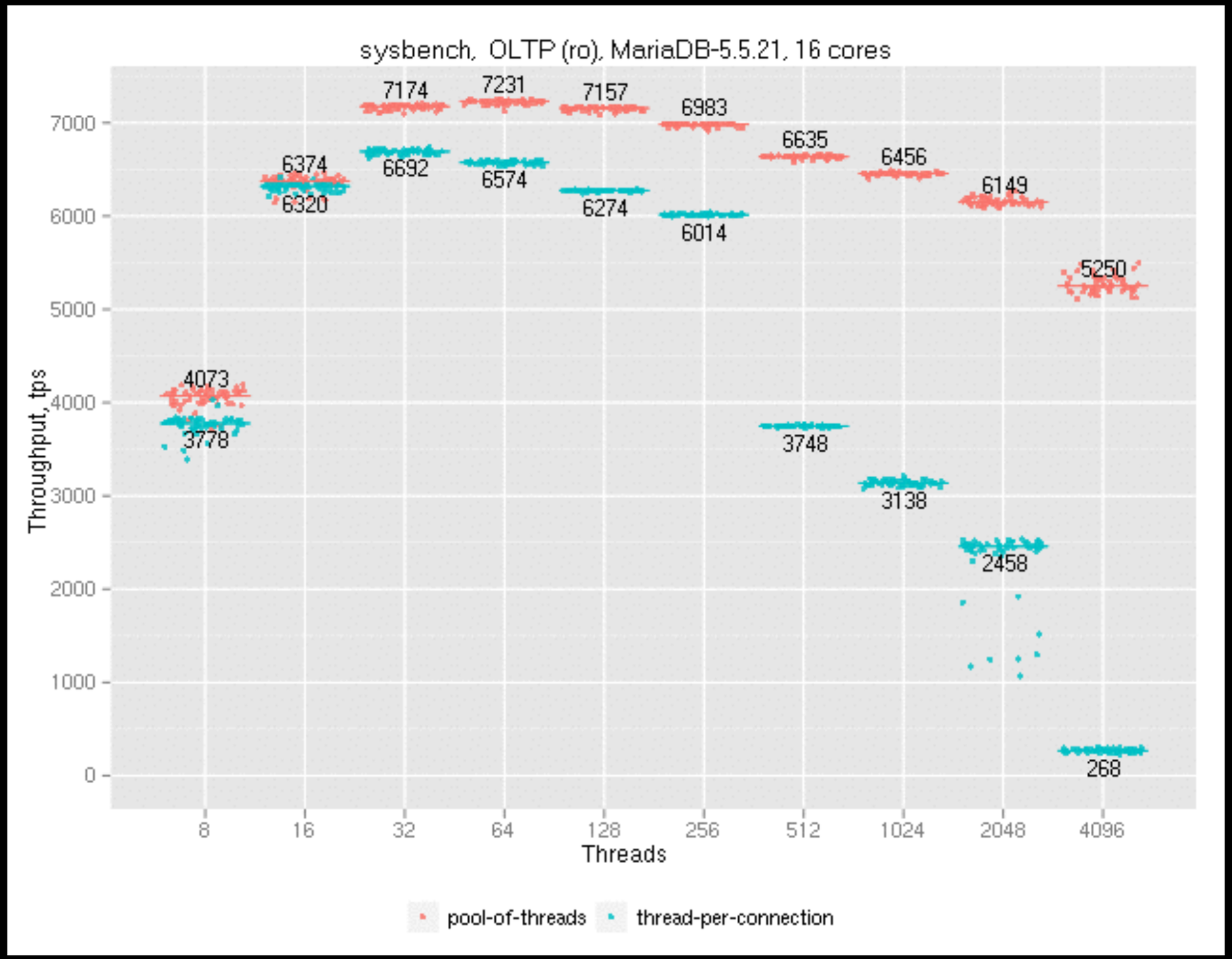
Replication: START TRANSACTION WITH CONSISTENT SNAPSHOT

- 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/

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

InnoDB improvements

- Multi-threaded flush (also in 5.7, different implementation + we're first)
- Page compression (optimised for Flash, SSD, FusionIO)
- 64KB pages in InnoDB (old limit = 16KB).
- Defragementation (FB, backported by DaumKakao)
- 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.

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;

Optimiser enhancements

- UNION ALL without temporary tables (5.7)
- Improve ORDER BY in optimiser
- Mostly there is 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/

EXPLAIN ANALYZE

```
MariaDB [information_schema]> explain format=json select * from all_plugins\G
*************************** 1. row ***************************
EXPLAIN: {
  "query_block": {
    "select_id": 1,
    "table": {
      "table_name": "all_plugins",
      "access_type": "ALL"
    }
  }
}
1 row in set (0.01 sec)
```

Passwords

- Password validation plugin exists now
  - https://mariadb.com/kb/en/mariadb/development/mariadb-internals-documentation/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’;
```

CONNECT

- CONNECT having full JSON/BSON support
- Can read filename.json files with ease
- Writing — INSERT, UPDATE, DELETE is supported — however, if you have added/modified values for objects or arrays, there can be complications (similar like the XML type object issue)
- Works with Oracle's JSON UDFs as well

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)
```

New 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]
```

Virtual columns

- A column in a table that has its value automatically calculated either with a pre-calculated/ deterministic expression or values of other fields in the table

- PERSISTENT (computed when data is inserted or stored in a table) or VIRTUAL (like a VIEW)

- Similar to MS SQL or Oracle

- https://kb.askmonty.org/en/virtual-columns/

GIS

- Full compliance for the OGC standards around GIS.
  - yes, we are missing a few functions, but its likely to improve
  - MDEV-4045 Missing OGC Spatial functions.
  - MDEV-60 Support for Spatial Reference systems for the GIS data.
  - MDEV-12 OpenGIS: create required tables: GeometryColumns, related views.
  - Speaking shortly, the MariaDB GIS part is now OpenGIS compliant, and passes all the OpenGIS conformance tests

Thank you!

Colin Charles
colin@mariadb.org / byte@bytebot.net
http://bytebot.net/blog | @bytebot on twitter
slides: slideshare.net/bytebot