本文是作者在ACMUG 2016 MySQL年会上的演讲内容，版权归作者所有。

中国MySQL用户组（China MySQL User Group）简称ACMUG。ACMUG是覆盖中国MySQL技术爱好者的一个技术社区，是Oracle User Group Community和MairaDB Foundation共同认可的MySQL技术社区。

我们关注MySQL，MariaDB，以及其他一切周边的开源数据库和开源工具，我们交流使用经验，推广开源技术，为开源贡献力量。

我们是开放社区，欢迎任何关注MySQL及其相关技术的人加入，我愿意跟其他任何技术组织和团体保持沟通和展开合作。

我们期望在我们的活动中大家都能以开心的、轻松的姿态交流技术，分享技术，形成一个良性循环，从而每个人都可以有一份收获。

ACMUG的口号：开源，开放，开心

关注ACMUG公众号，参与社区活动，交流开源技术，分享学习心得，一起共同进步。

MyRocks

Space and write optimized OLTP database at Facebook

Yoshinori Matsunobu
Production Engineer, Facebook

1 Our main MySQL database

2 Issues in InnoDB/B+Tree database

3 MyRocks overview and features

4 Benchmarks

5 Migrating from InnoDB to MyRocks in produc

6 Future works

Storing social graphs

Sharded MySQL database

Petabytes scale

Low latency serving queries

Automated operations

H/W trends and limitations

	HDD	Flash
Cost per GB	Low	High
Current Capacity	4-10 TB per drive	1-4 TB per drive
Read IOPS	>100	< 20000
Write IOPS	>100	< 10000
Write Endurance	Unlimited	Limited

Space and write inefficiency in InnoDB

Single row modification results in entire page write

RocksDB

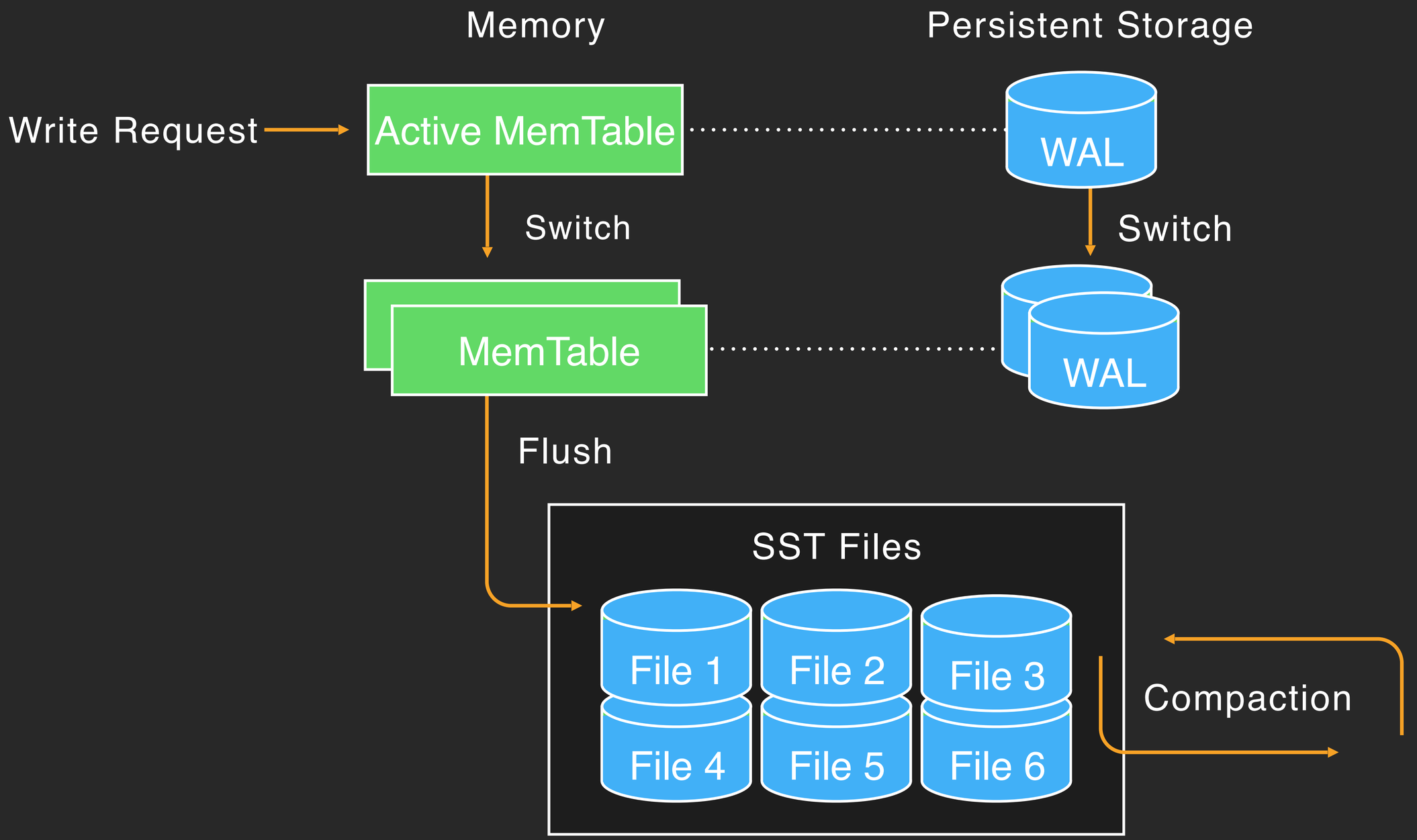
Open source LSM database

Forked from LevelDB

Key-Value LSM persistent store

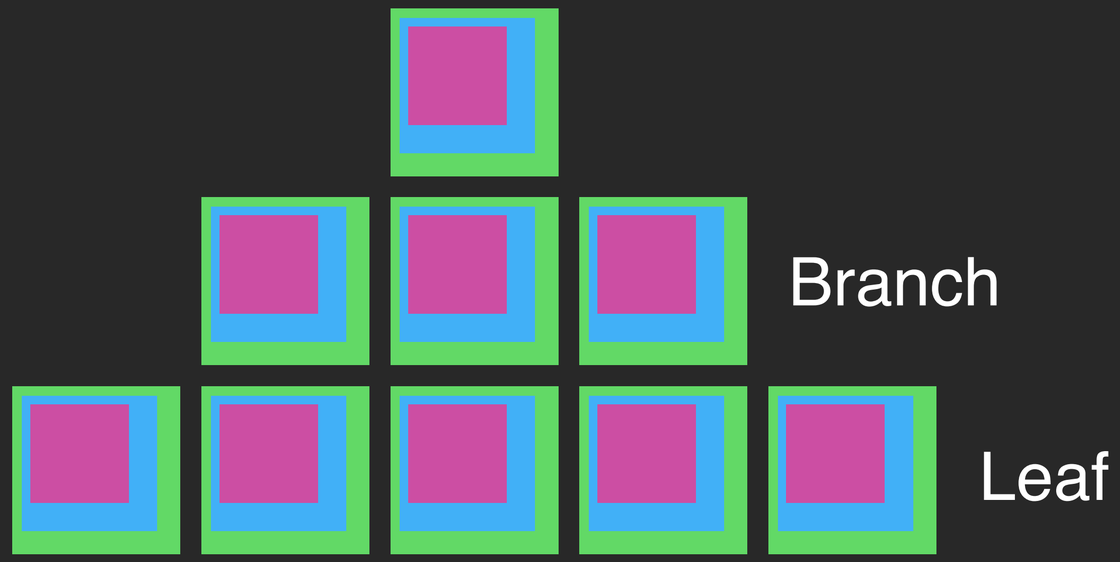
Easier integration

Native compression

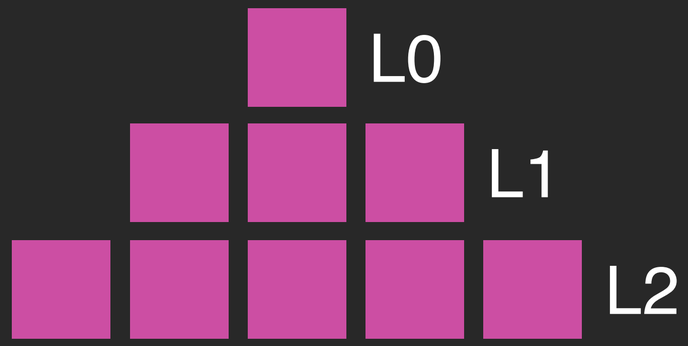

Memory
Persistent Storage
Write Request
Active MemTable
WAL
Switch
Switch
MemTable
WAL
Flush
SST Files
File 1
File 2
File 3
File 4
File 5
File 6
Compaction

Space differences between InnoDB & RocksDB

InnoDB

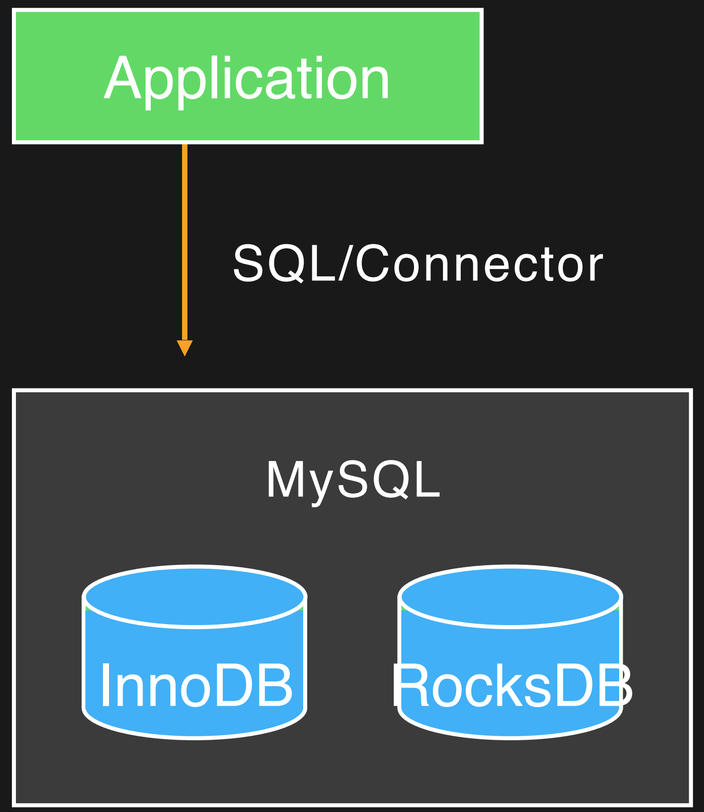


RocksDB



MyRocks

RocksDB storage engine for MySQL



Taking both LSM advantages
and MySQL features

Fully Open Source

MySQL features

Easy to access

Major features in MyRocks

Similar feature sets as InnoDB	Transactions	Online Backup
	Atomicity Non locking consistent reads Read Committed Repeatable Read Crash safe slave and master	Logical backup by mysqldump Binary backup by myrocks_hotbackup

Performance

LinkBench

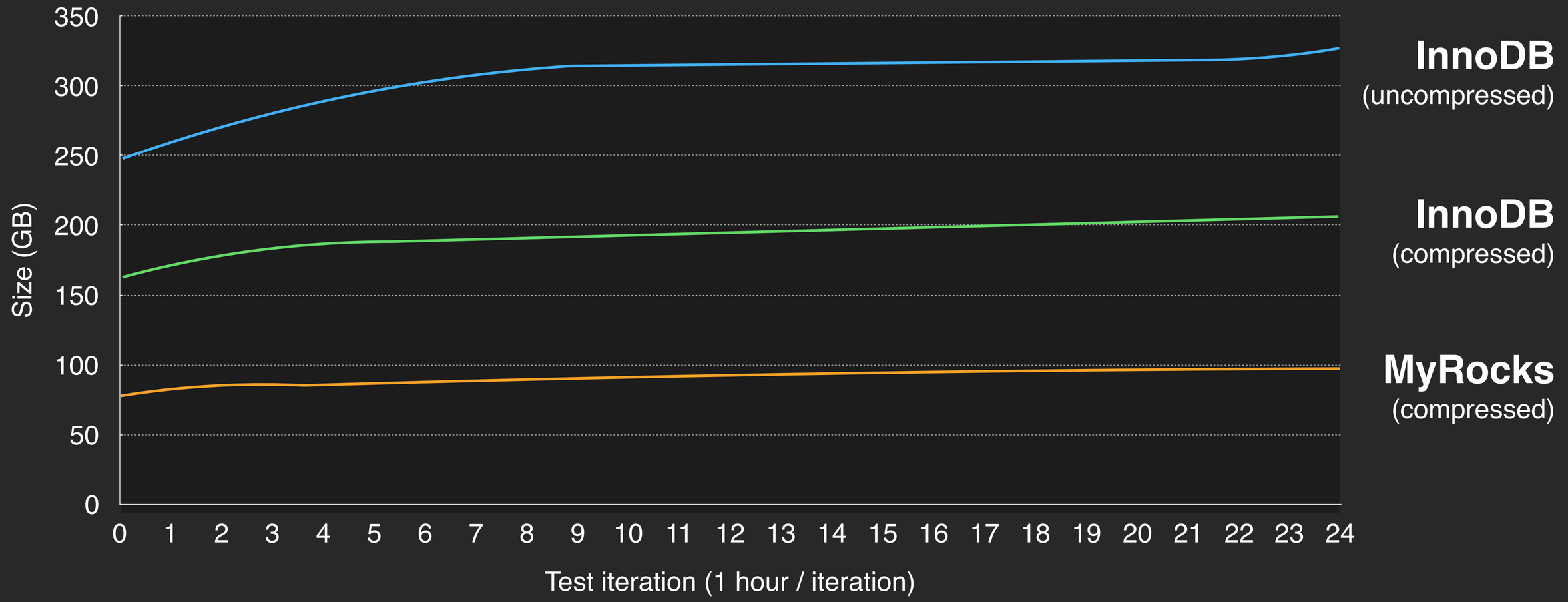
Space Usage

TPS on Flash & Disk

Flash writes per query

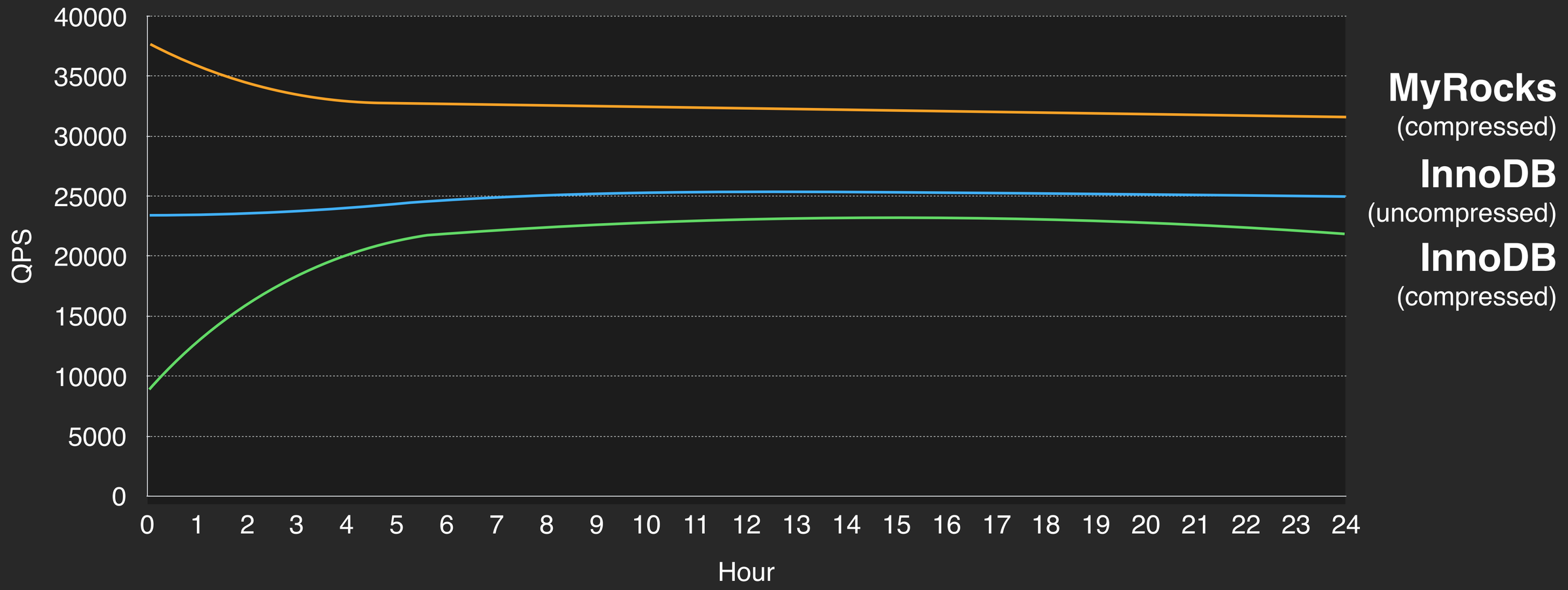
code.facebook.com

Database size



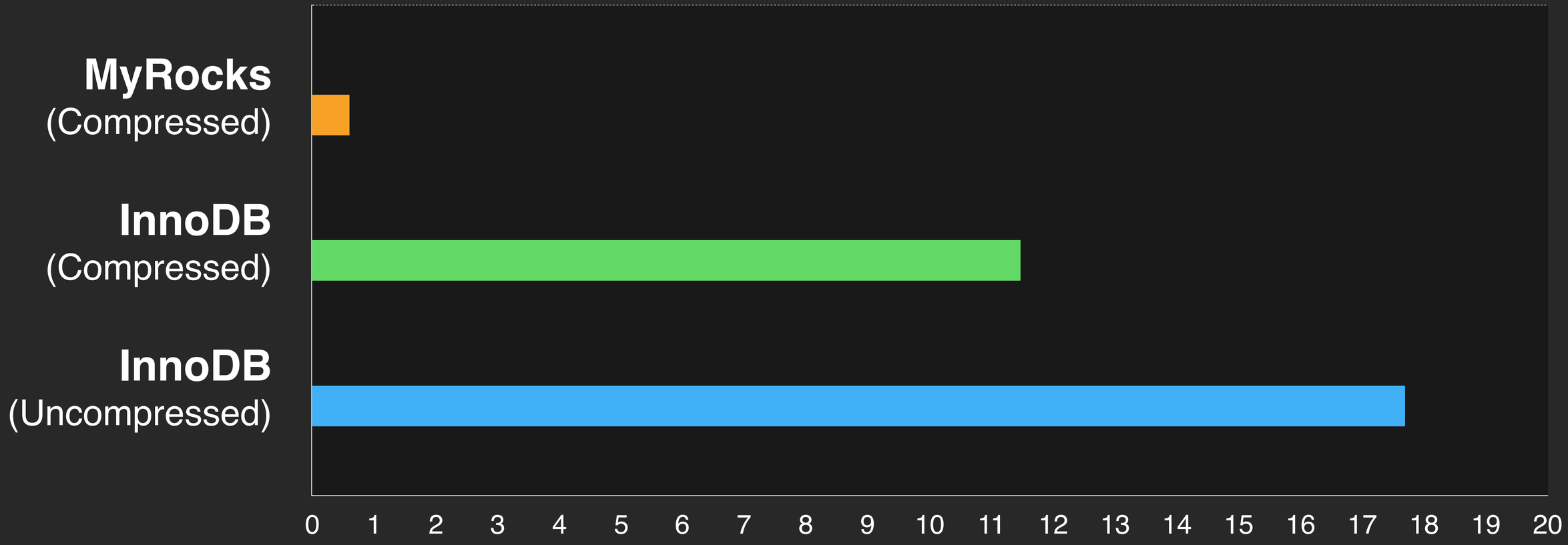
Transactions per second

Query rate, SSD, 20 clients, 24 hours



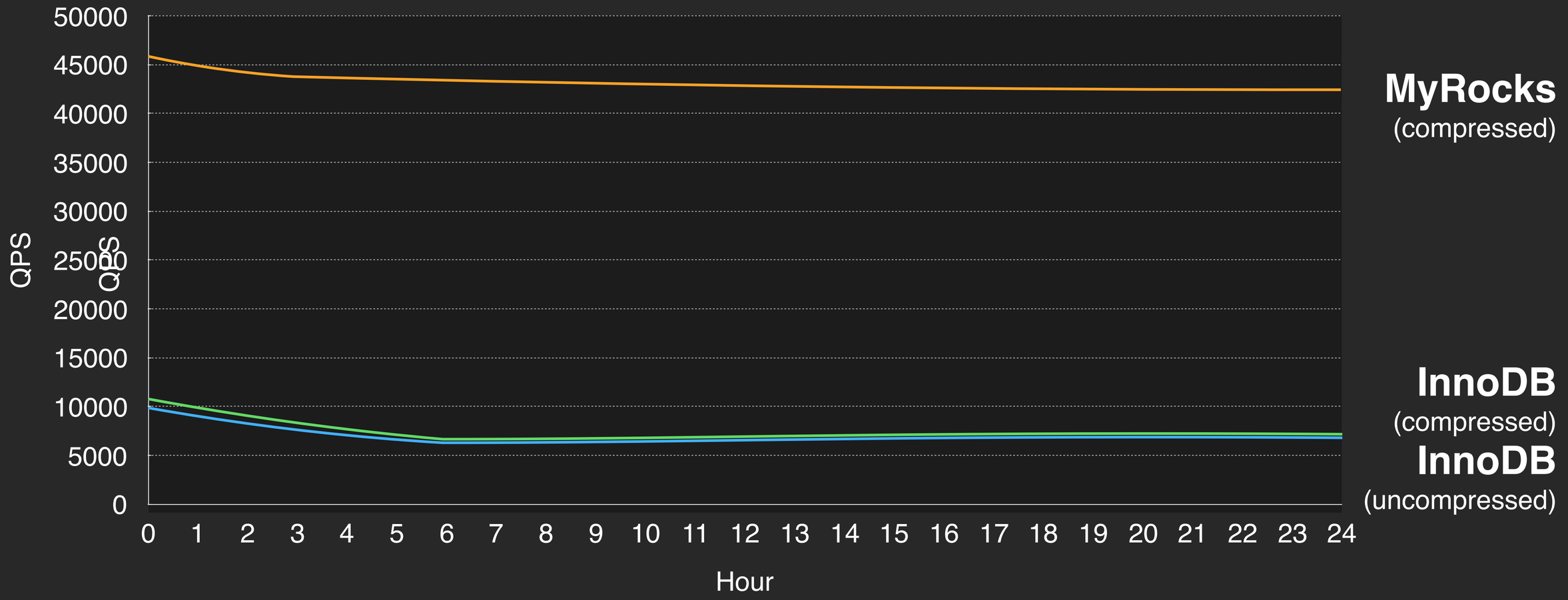
Write amplification

Relative KB written per query



On HDD workloads

Query rate, Disk, maxid1=10m, 20 clients, 24 hours



Migrating from InnoDB to MyRocks

Create MyRocks Instances without downtime

Create MyRocks instances within reasonable time

Online data verification

5% production on main MySQL database in one of our regions

Migrating from InnoDB to MyRocks

50%

reduction in MySQL storage reduction

Future work

Increasing our MyRocks deployment in production

More features for external use cases

- Two phase commit
- Online DDL
- Foreign Keys

More documentation

MyRocks

Open source:

https://github.com/facebook/mysql-5.6

More info at:

code.facebook.com