使用TerarkDB提升MySQL的性能和压缩率

主讲人：Terark 联合创始人 郭宽宽

互联网和大数据带来的挑战

海量用户带来的海量随机访问

- 写入性能目前通过单节点的持续写入，多节点只读，一般能满足需求（如阿里云的 PolarDB ）
- 绝大多数互联网用户对在线服务的访问偏向于随机读
  - 比如新闻资讯、搜索引擎、舆情监控、电商类商品检索等
  - 对大量随机读的优化，目前没有很好的解决方案(目前只能增加内存或建立额外的索引缓解)
- 非随机读的场景，往往允许离线进行处理，一般不要求事务，暂时还可以忍

内存和SSD依然很贵

- 公司每个月的数据增长没有几个T，都不好意思跟人打招呼
- SSD 的价格依然是机械硬盘的数倍，容量越大，价格差距越大
- SSD 的寿命非常有限

数据库领域的探索没有止境

数据库

MySQL | MongoDB | TiDB | CockroachDB | Cassandra

PostgreSQL | MariaDB | PolarDB | HBase

存储引擎

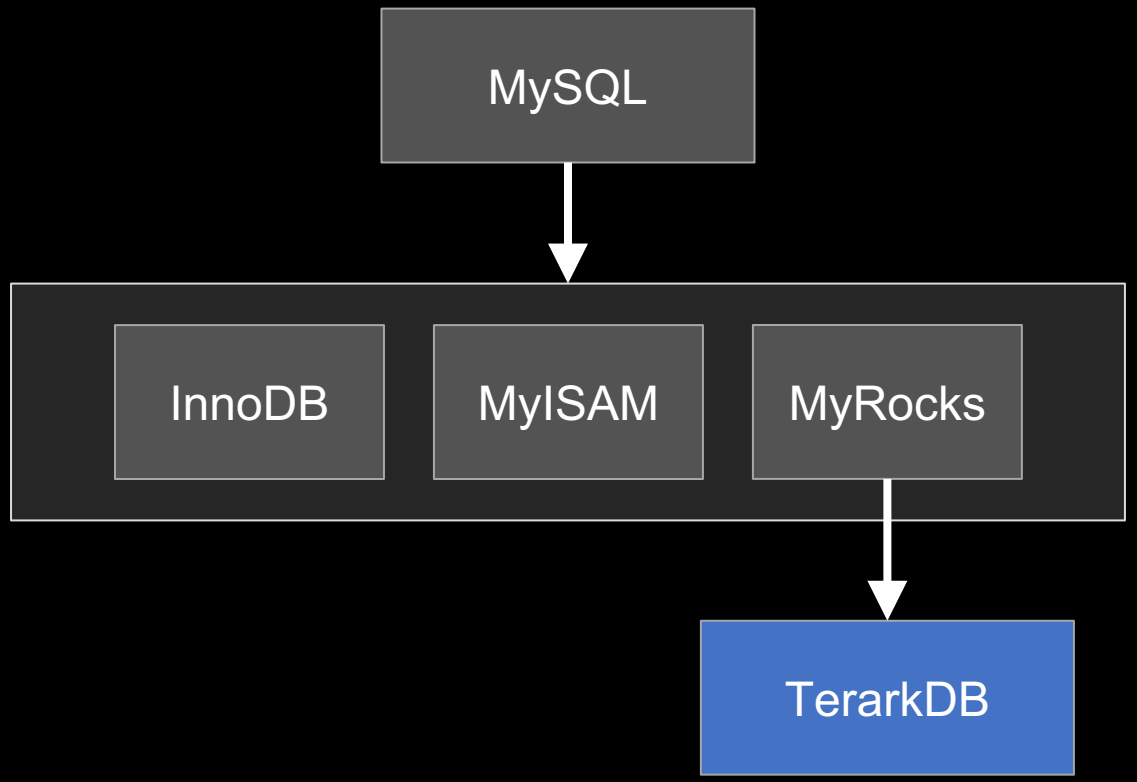
TerarkDB | InnoDB | RocksDB | LevelDB | WiredTiger

为什么从引擎层优化

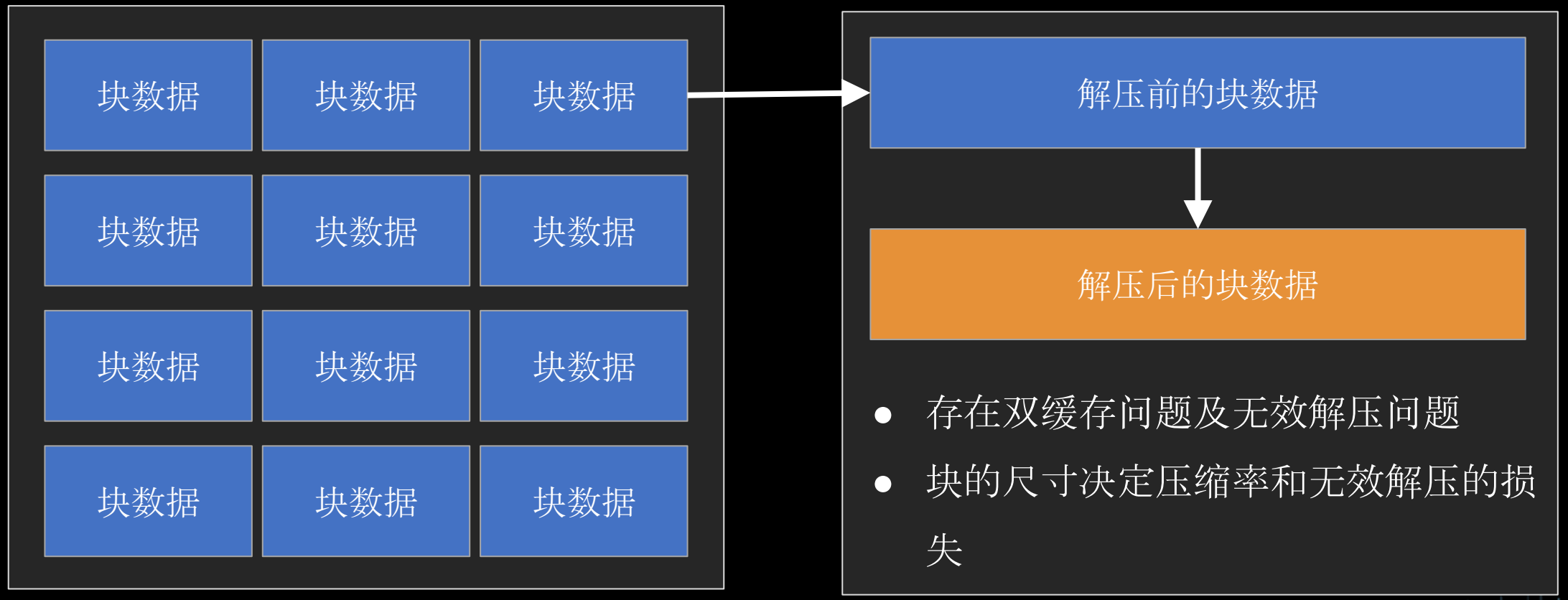
- 目前的存储引擎底层算法，针对随机读较多、内存受限的场景，还有很大的改进空间
  - 块压缩对随机访问很不友好
  - 压缩率太低
- 引擎层的修改不触及用户现有数据库的使用逻辑，更加透明
- 引擎层更加通用，可以适应各类数据库产品，应用范围更广，也能和现有的数据库产品充分整合，设计各类解决方案

TerarkDB 对 MySQL 的改进方式

- MySQL 以 RocksDB 作为存储引擎是发挥了 RocksDB 的 LSM 随机写速度快的优势，也是 Facebook 目前使用的主要方法(MySQL on RocksDB, 简称 MyRocks)
- TerarkDB 基于 RocksDB 的接口，将自己的算法适配到了 MyRocks 中，进一步支持了 MySQL 数据库



“块压缩”的问题



- 存在双缓存问题及无效解压问题
- 块的尺寸决定压缩率和无效解压的损失

Terark 可检索压缩算法

全局压缩

不需要解压即可提取单条记录

单条目标记录

- 不需要常见的数据库使用的 DB Cache 层来缓存解压后的块
- 压缩率逼近理论上限（相当于块压缩中的块就是整个库）

Terark 可检索压缩算法

Terark 的可检索压缩算法，由以下两部分组成：

- 索引压缩算法：CO-Index（Compressed Ordered Index）
  - 树结构高度压缩
  - 具有通过 ID 反查 KEY 的功能（区别于传统B+树）
- 数据压缩算法：PA-Zip（Point Accessible Zip）
  - 全局压缩
  - 提取单条数据时无需多余解压

索引压缩算法：CO-Index

Succinct Data Structure

- Succinct 数据结构历史悠久，但是并未引起大家的重视，对于相同的树结构，对比基于指针技术，它仅仅需要 1/30 的内存
- 使用位向量来表达树结构，开源实现有 Succinct Data Structure Library，缺点是性能较指针更低，需要通过工程上大幅度优化来接近指针性能

Nested Patricia Trie

- 原生的 Patricia 支持路径压缩，通过把一串仅包含一个孩子的节点，压缩成一个包含多个字符的节点
- 我们对其进行了更进一步的嵌套压缩：把压缩后的路径构建成一个全新的 Patricia Trie，进一步提升数据的压缩率。

数据压缩算法：PA-Zip

LZ系列算法的变种

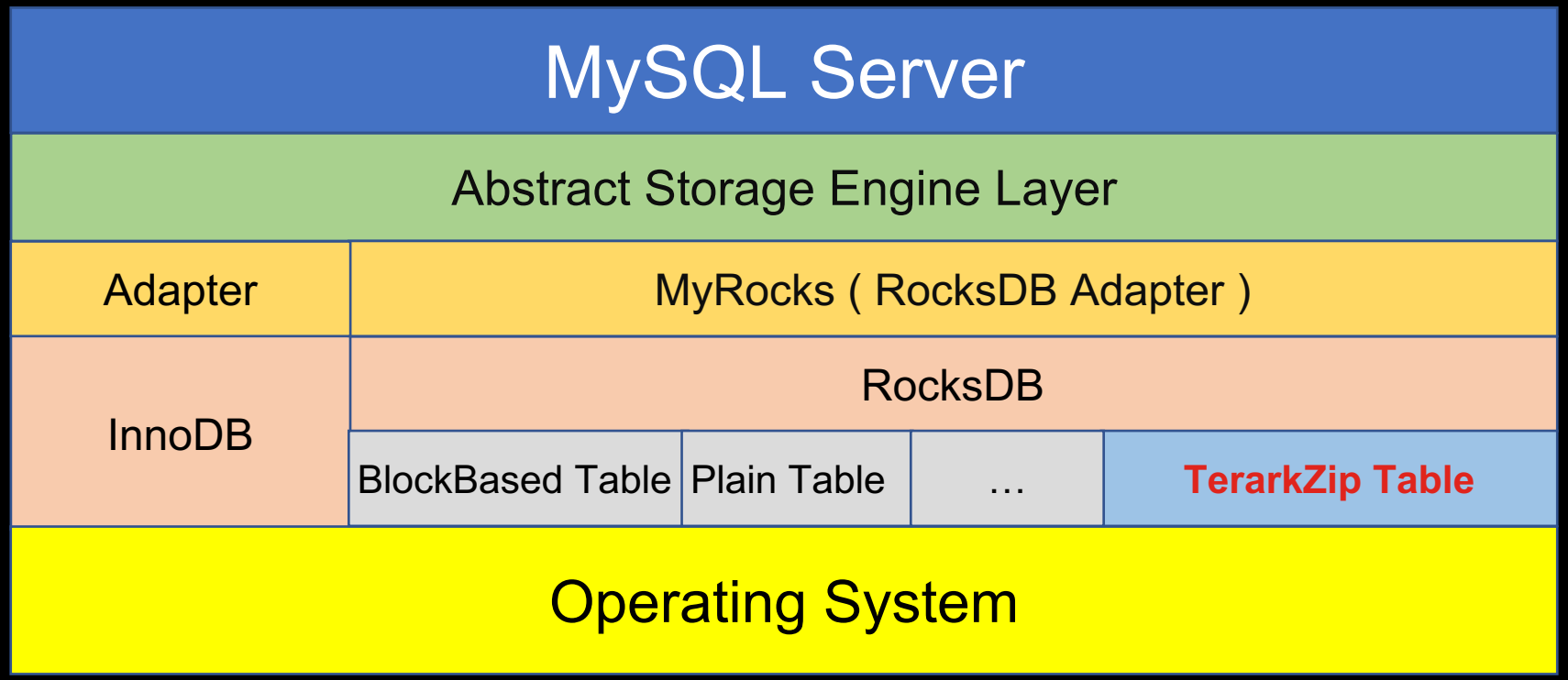
- 基于 LZ 系列算法进行了大幅度的改进，使用“全局字典” + “局部字典”的方式，将压缩率达到最理想的程度
- 采用滑动窗口的方式进行数据压缩
- 根据实际测试，全局字典的尺寸限制为 12GB 以内，效果比较理想，更大的字典对压缩率的帮助很有限

数据压缩算法的缺点

- 由于需要进行大量的计算，在数据写入过程中对 CPU 的消耗会比较高
  - 目前可以采用写入限流的方式减轻 CPU 负载（大多数情况下，并不需要全速写入）
  - 在大型系统架构下，可以采用计算和存储分离的逻辑，单独进行数据压缩

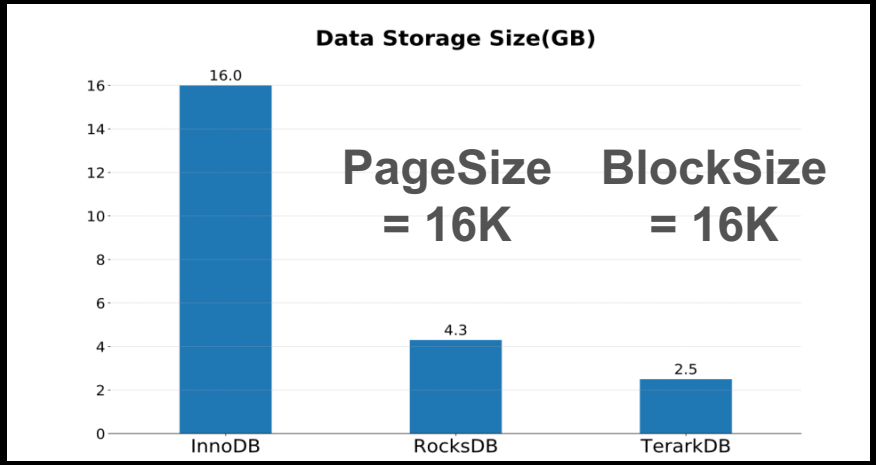
集成进入 MySQL 数据库

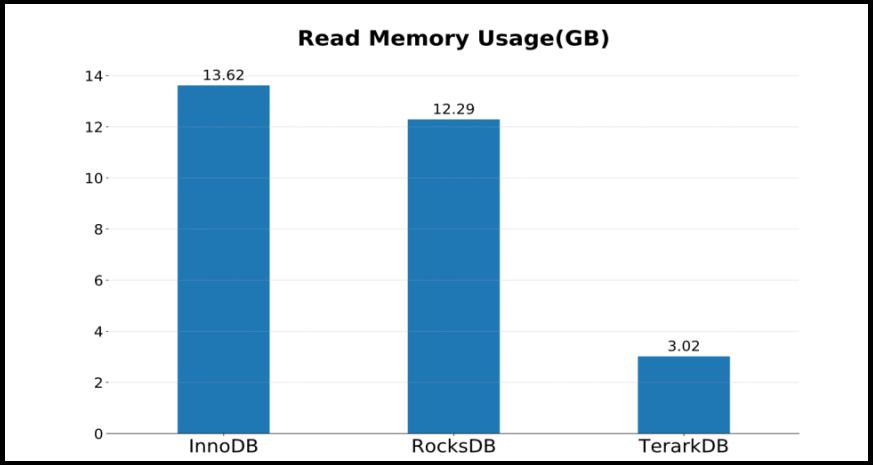
TerarkDB 通过 Facebook 推出的 MyRocks 适配进入 MySQL，整合了 RocksDB 本身的优秀调度层和 TerarkDB 的底层算法：

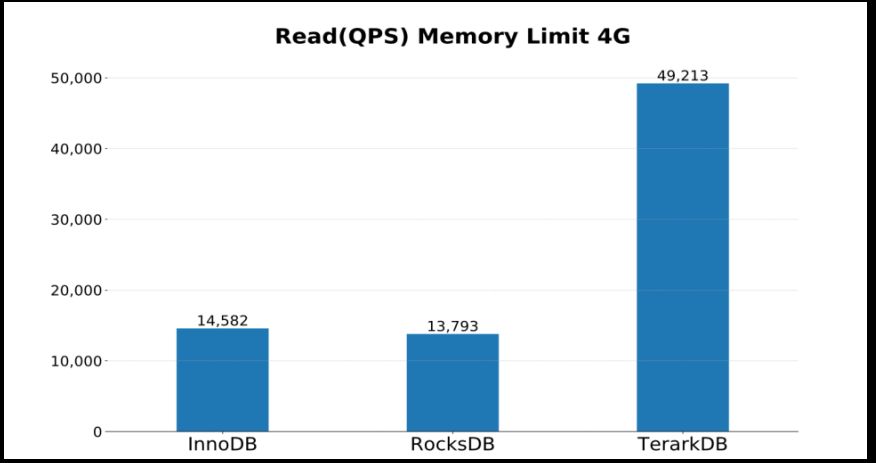


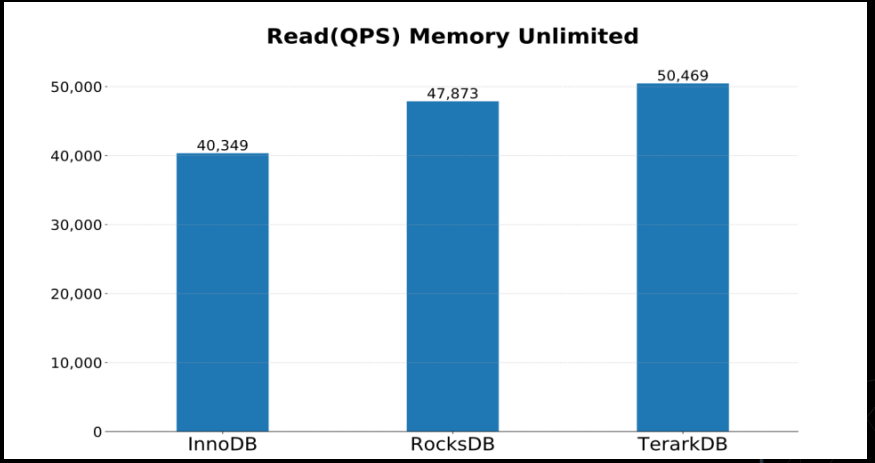
性能和压缩率对比

Amazon Movie Reviews Open Dataset, 原始数据 9.1GB









谢谢各位！

- 目前可以通过官网直接下载试用
- 和 MyRocks 100% 兼容
  - MyRocks 对 MySQL 有部分功能限制，如不支持外键约束等
- 官方网站：www.terark.com
- 微信公众号：TerarkLab