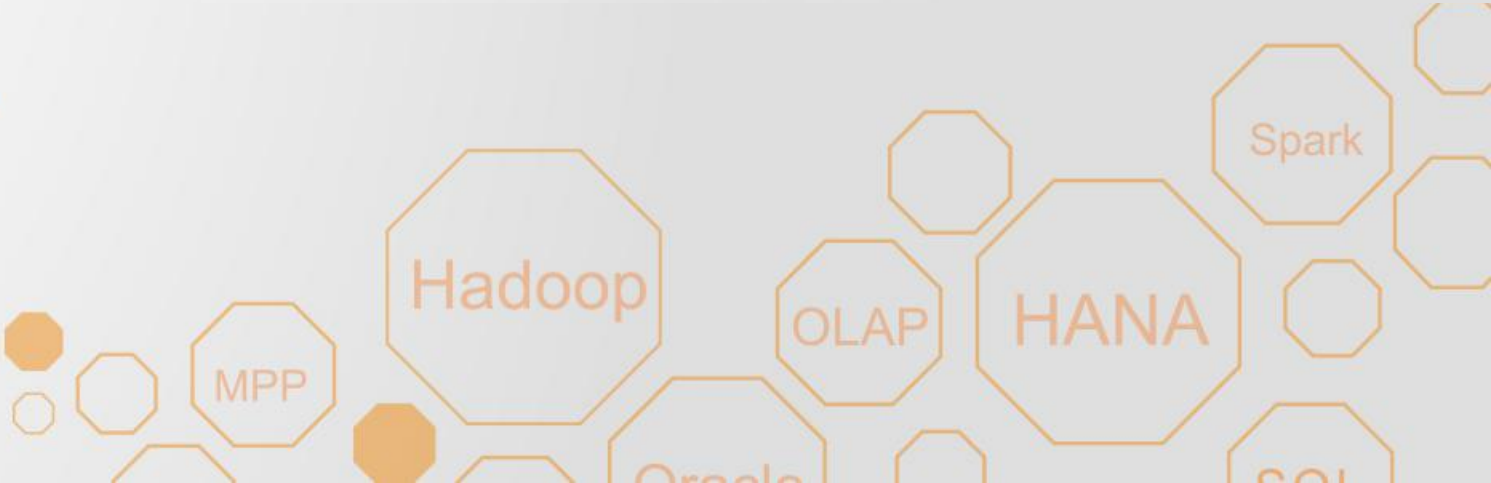
腾讯云原生自研数据库内核深度优化最佳实践

李昕龙　腾讯云数据库高级工程师



目录

TDSQL-C架构介绍

产品的背景和架构主要特性

产品背景

	存储容量	可靠性	可用性	可扩展性
传统架构问题	1、磁盘容量有限 2、扩容对业务影响大 3、分库分表对业务影响大，分布式事务问题多	1、普通复制（binlog）可能丢失数据（RPO>0） 2、同步复制性能差	1、HA、恢复速度慢（RTO分钟级） 2、副本时延大（分钟级-小时级）	1、水平扩展需要完整数据库副本，产生大量IO 2、只读副本部署速度慢（分钟级-小时级）
用户需求	1、大于100T容量 2、快速、透明扩容	1、不能丢失数据（RPO=0） 2、多副本容灾	1、快速HA、恢复（RTO秒级）、回档 2、更小的副本时延（毫秒级）	1、秒级副本扩展
技术方案	1、云存储：理论无上限，多副本可靠性，持续备份，归档等		1、数据分片：并行恢复和回档 2、物理复制：页面粒度并行复制	1、共享存储：减少大量冗余IO

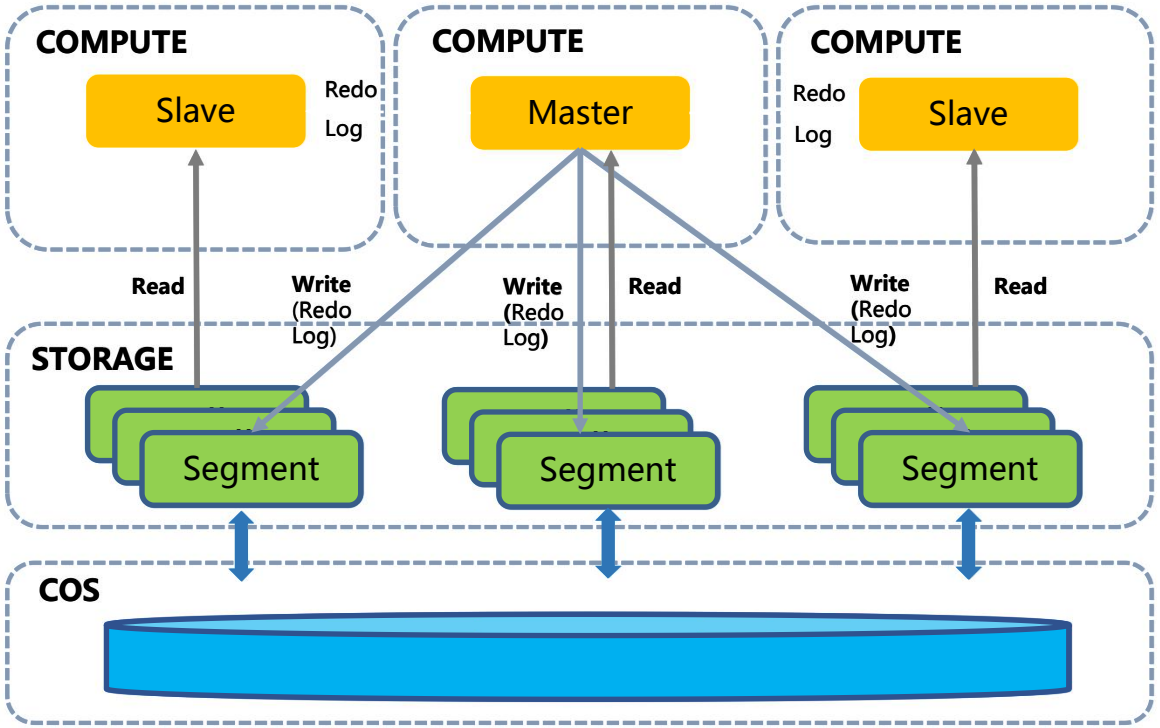
架构特性

1PB

海量存储自动扩容

96C 768GiB

100W QPS



100%兼容MySQL

秒级 扩展15个只读节点

毫秒级 只读延时

秒级 故障切换

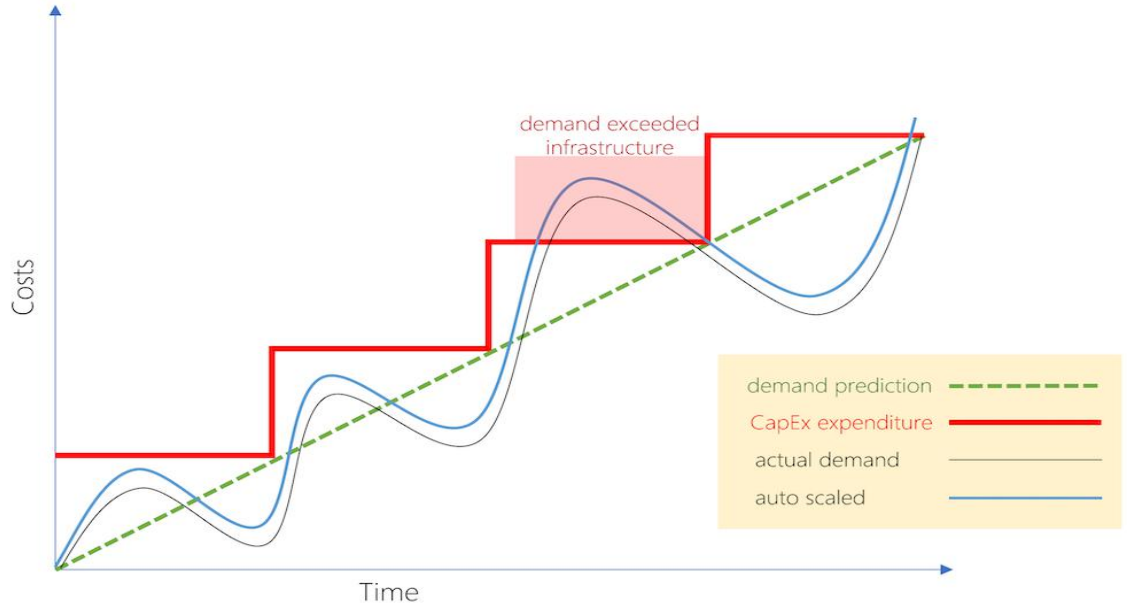
秒级 快照备份

Serverless

用户场景实践

典型客户场景分析

Serverless



场景

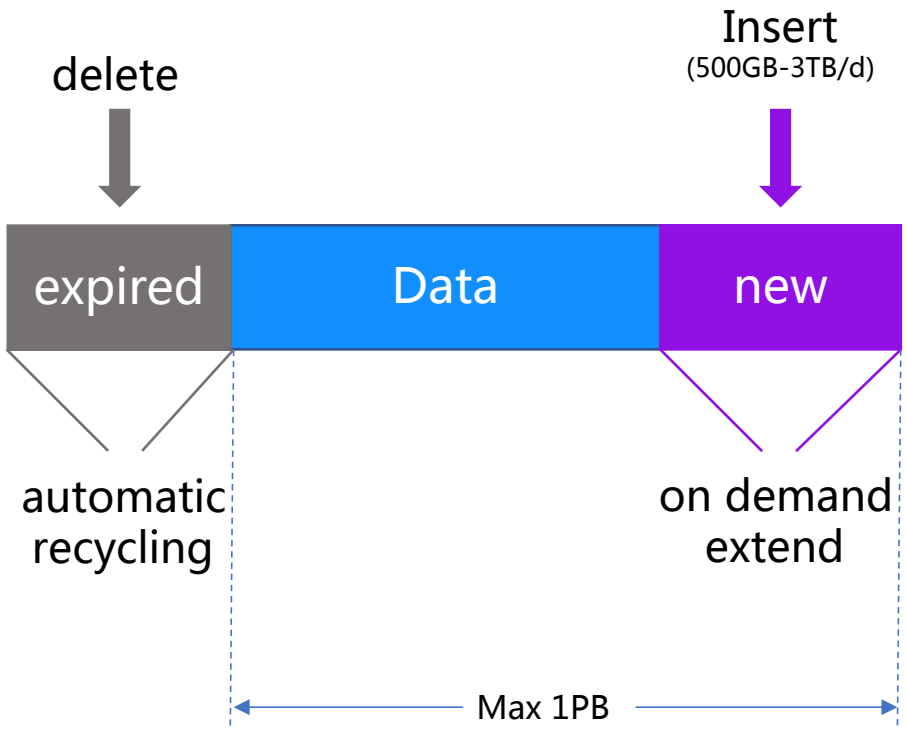
- 开发测试场景，低频使用数据库
- IoT，边缘计算，SaaS平台，负载变化频繁

特性

智能极致弹性：极速启停，根据负载启停实例

按需计费： 按实际使用的计算和存储量计费，不用不付费；按秒计量，按小时结算

弹性容量



场景

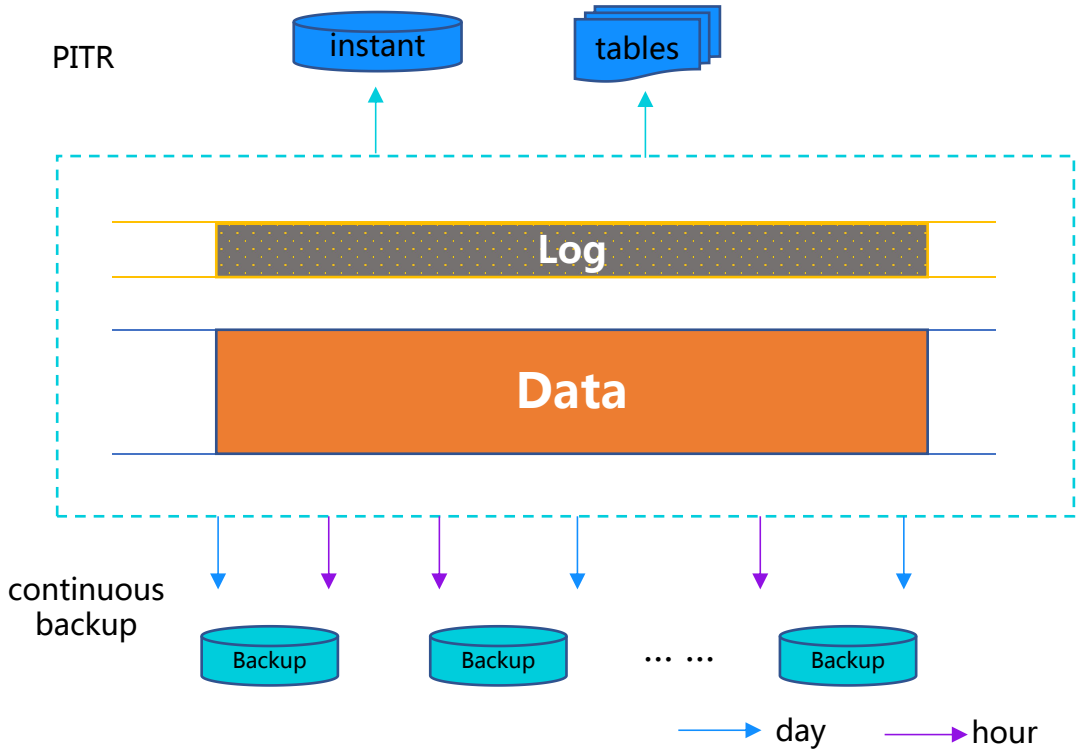
- 新数据产生速度非常快，对单库容量要求高
- 开发测试场景，数据生命周期短
- 历史数据存储，用户关注空间扩缩容成本

特性

按需扩容：存储根据操作页面按需扩容，无需提前扩展空间

自动回收：空闲空间自动回收，按实际使用容量计费

备份回档



场景

- 金融行业相关场景，对备份的时效和速度有较高的要求
- 游戏业务，频繁回档数据，对备份回档速度要求高
- 用户可能存在的误操作，对快速回档要求高

特性

持续备份：存储分片根据备份点进行独立备份，同时做到备份全局一致性备份

并行回档：每个分片并行回档数据全量/增量备份，并行回放日志

PITR：库表快速回档到任意时间点

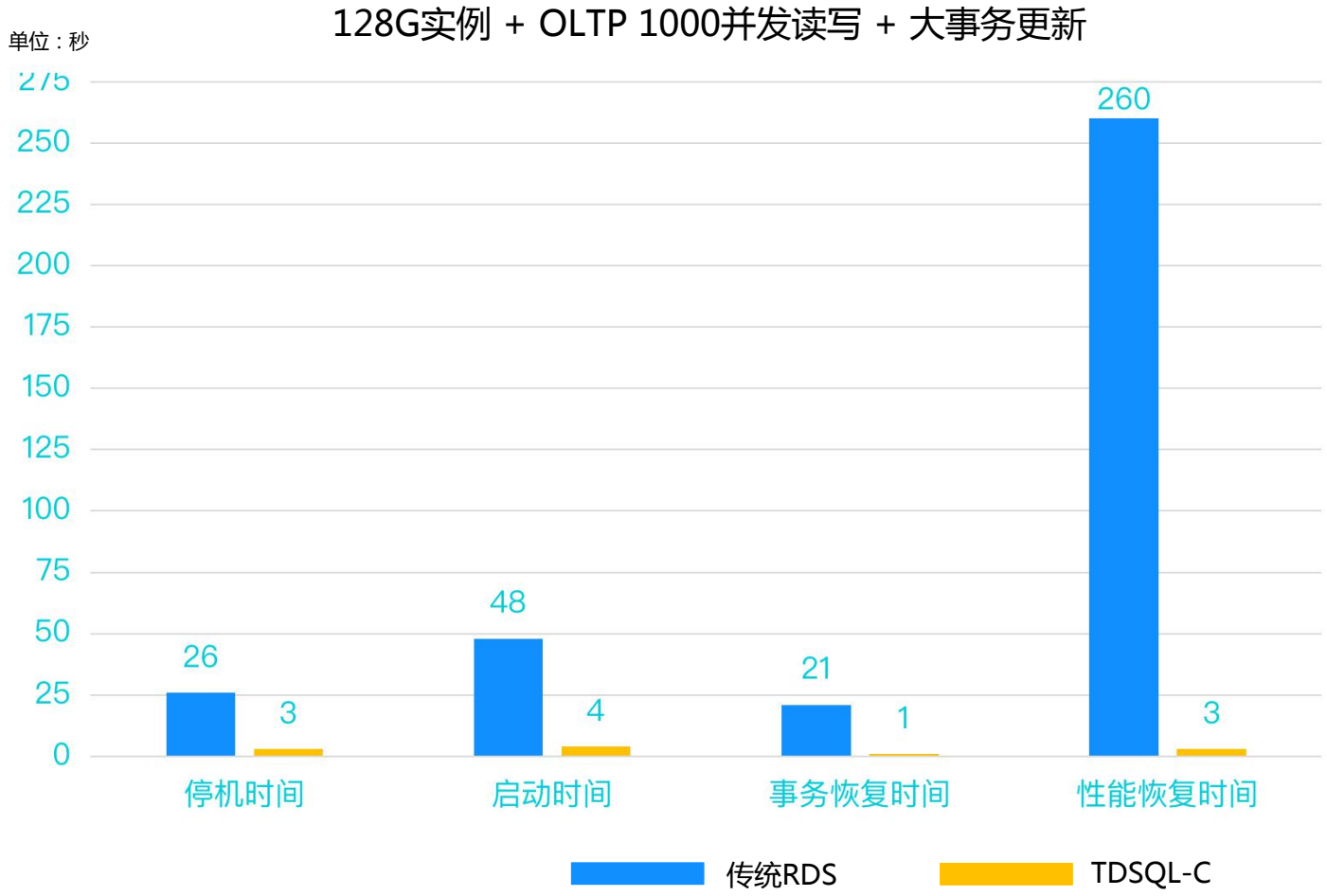
系统关键优化

支撑相关场景的特性优化

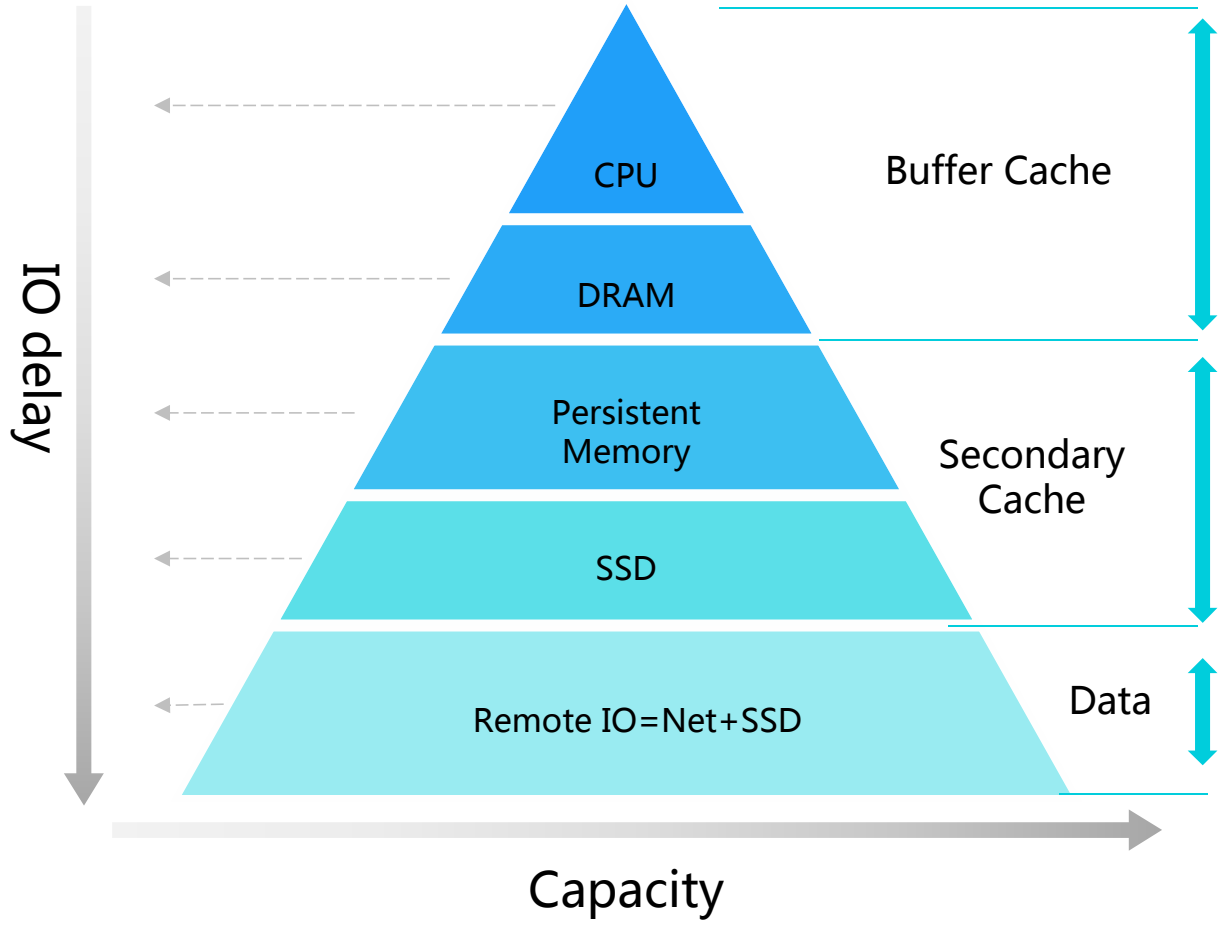
极速启停

优化

- BP并行初始化
- 独立BP
- 事务系统并行初始化
- 表锁恢复优化
- 并行恢复
- 快速停机



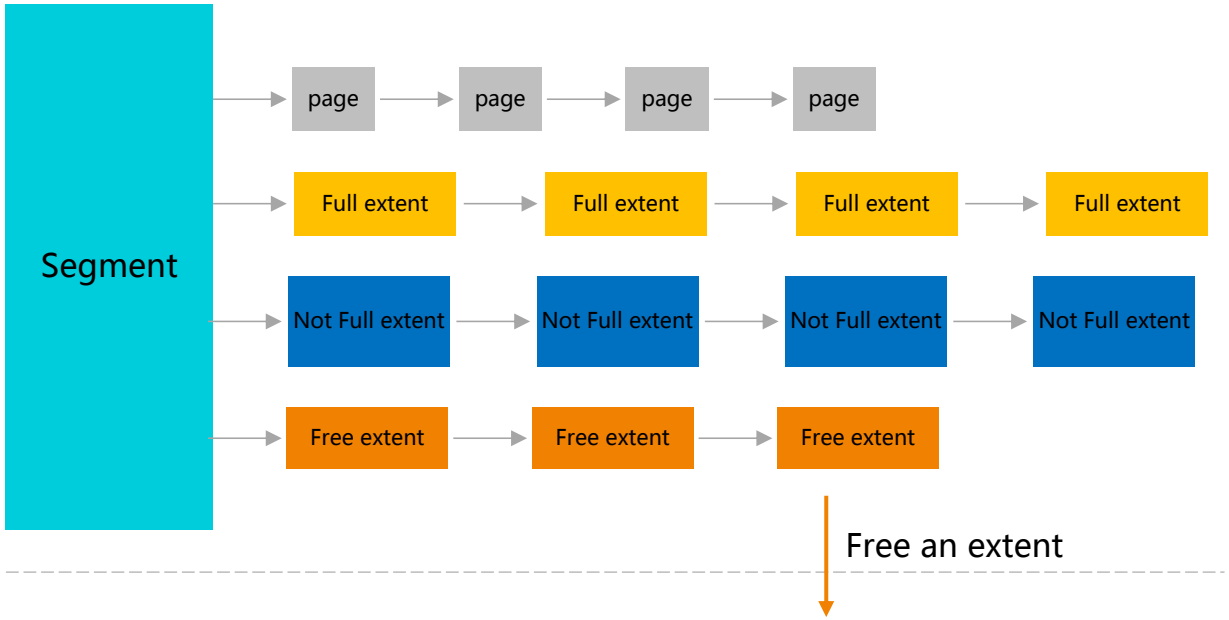
二级缓存



二级缓存：针对普遍存在的IO Bound场景，在计算层引入独立于Buffer Pool的二级缓存，利用非易失存储等新硬件的能力对BP进行扩容，缓存热数据，提供快速高效的热数据访问能力

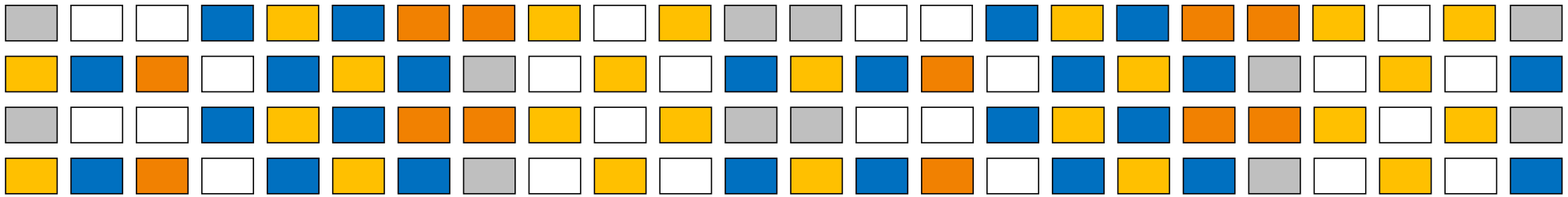
优化效果：随着数据量的增大，性能平均提升100%以上

极致伸缩



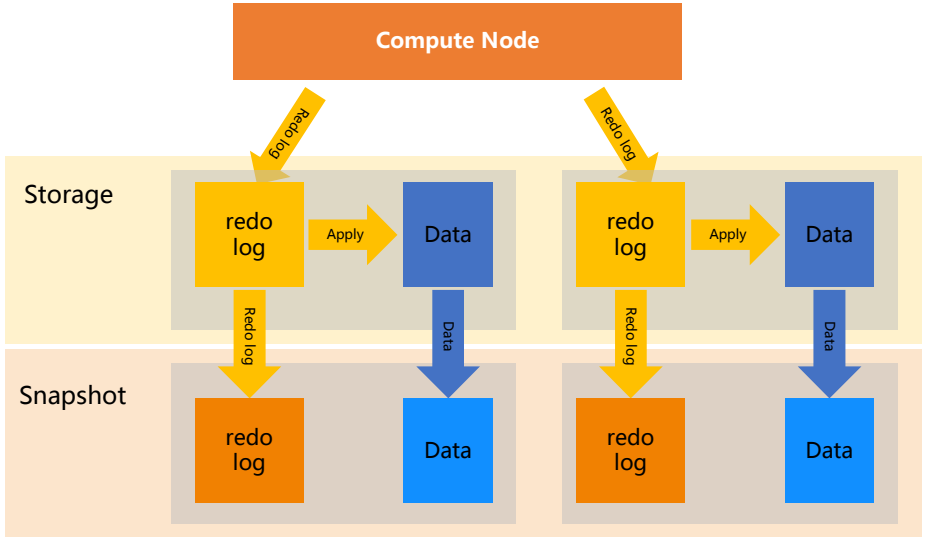
- 卸载空间扩展，日志驱动按需扩展
- 段管理以1M的extent为最小单位
- 存储池物理分配单元为1M
- 段空闲链表中extent定期触发自动回收
- 提供真正意义上的按需计费能力

Storage Pool

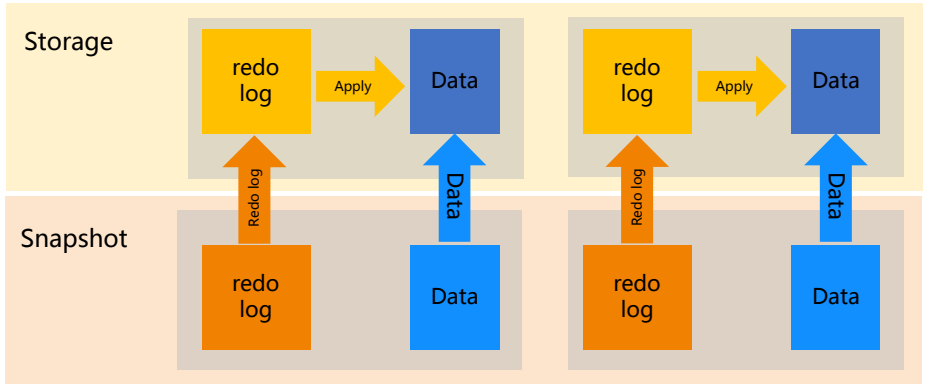


极速备份回档

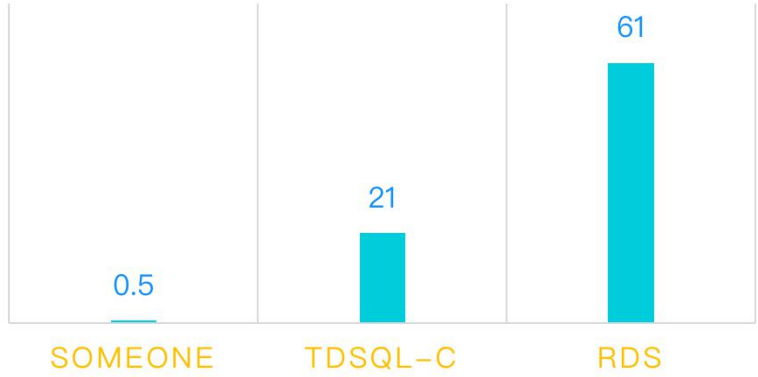
自治备份



并行回档

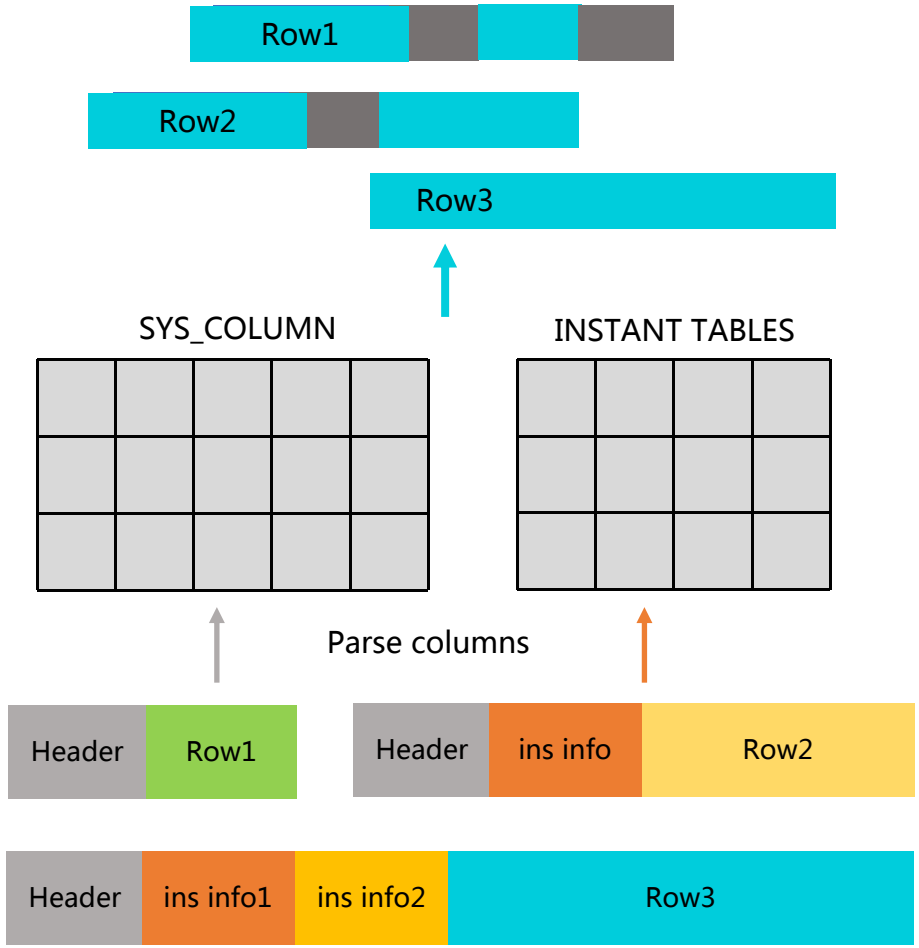


1TB 备份时间（单位：分钟）



1TB 回档恢复时间（单位：分钟）

Instant DDL



Instant DDL (O1)

- 新增列
- 修改列类型
- 删除列

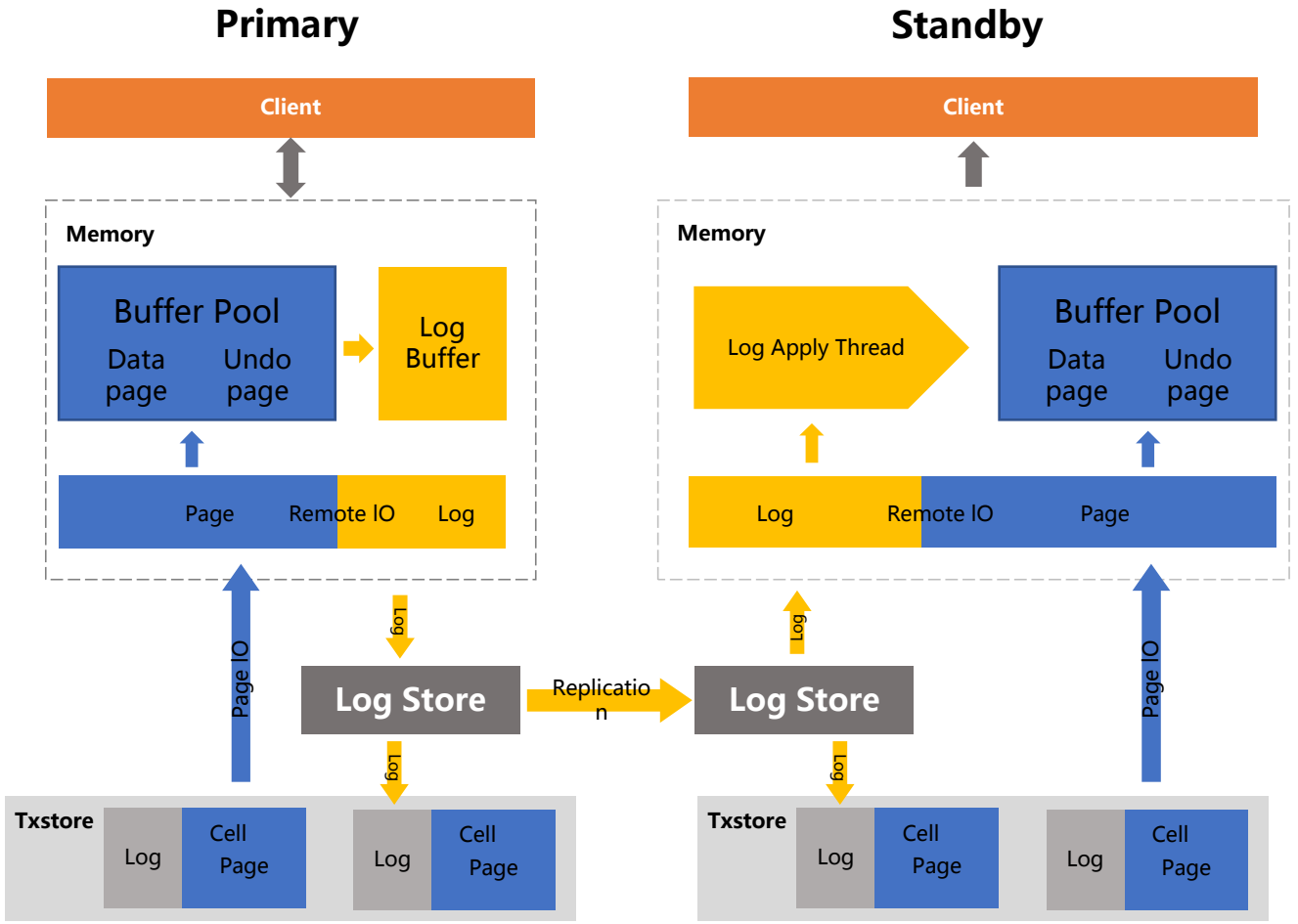
并行rebuild (提升200%)

- 并行扫描
- 并行导入

产品未来演进

用户为本，深度探索和优化

Global database

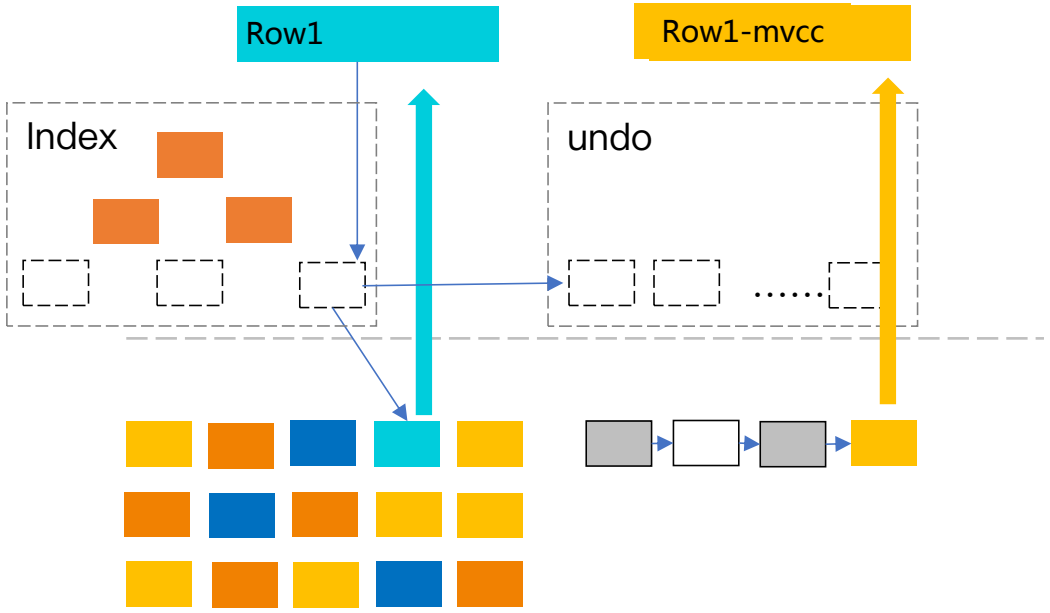


极致性能：Log store提升日志响应速度和整体吞吐量，提供极致的写性能

跨region读：提供可用性更高的、跨region的只读服务

金融级可靠性：跨region灾备，打造超高级别的数据可靠

计算下推

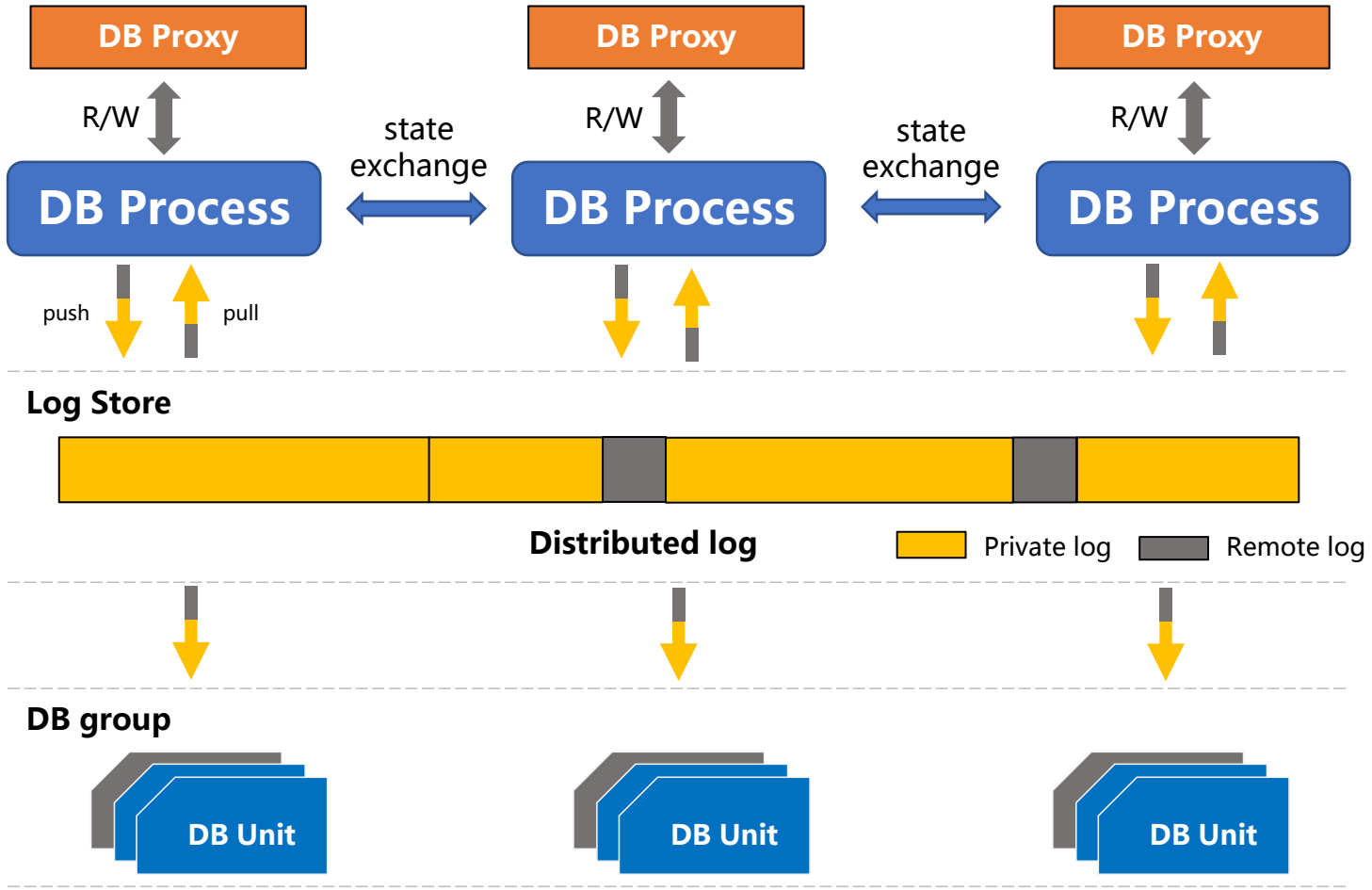


叶子结点读下推：将叶内扫描计算下推倒到存储完成，计算节点获得业内扫描结果

undo页面读下推:对于数据的历史版本扫描下推到存储层完成，内存不读取页面，仅获取中间结果和最终结果

写下推：特定页面的修改直接在存储层进行

多写架构



- 数据集分区
- 多节点读写
- 日志传输
- 全局事务

扫码关注

“腾讯云数据库”

体验移动端运维数据库

获取更多资讯


THANKS
时序
开源
自研
分布式
文档
Spark
HANA
SQL
OLAP
Oracle
Hadoop
MPP
DB2