
SACC
2021中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2021
数字转型 架构重塑
IT168.com
ChinaUnix.net
ITPUB
云上会议 网络直播丨 2021.5.20-2021.5.22

腾讯云原生数据库架构探索与实践

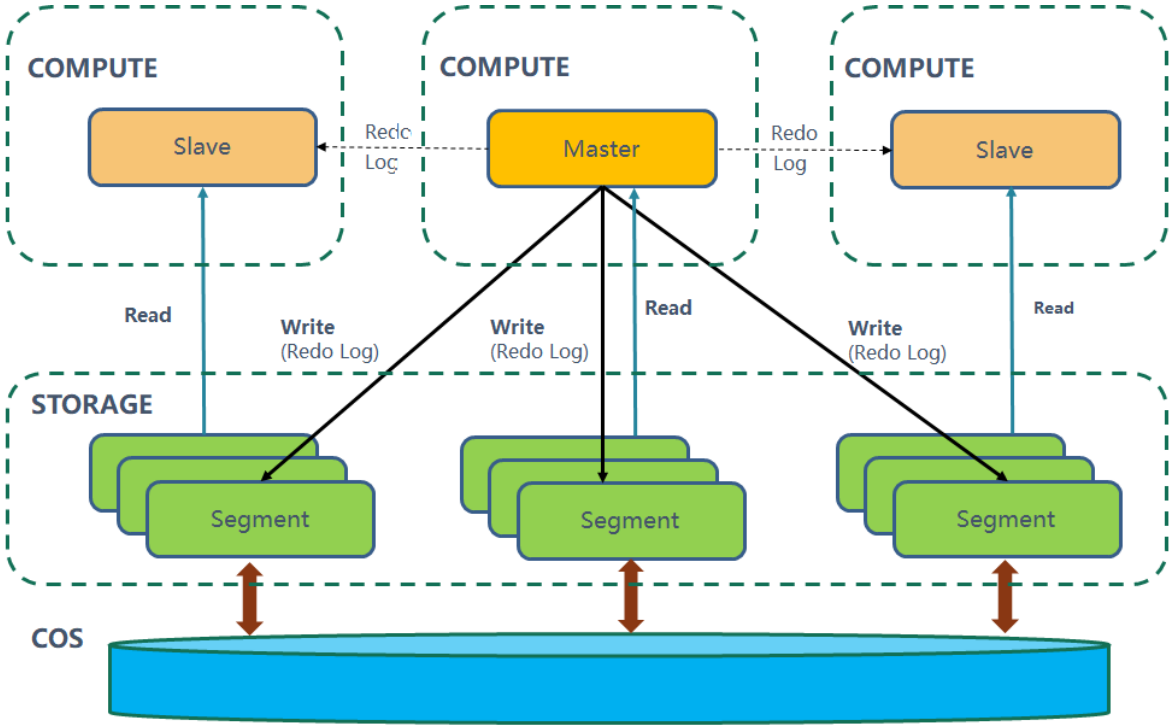
目录

概览：架构与特性

128TB
海量存储自动扩容

96C 768GiB
0.25C 0.5GiB

100W QPS



MySQL PostgreSQL
100%兼容多种引擎

秒级 扩展15个只读节点

毫秒级 只读延时

秒级 故障切换

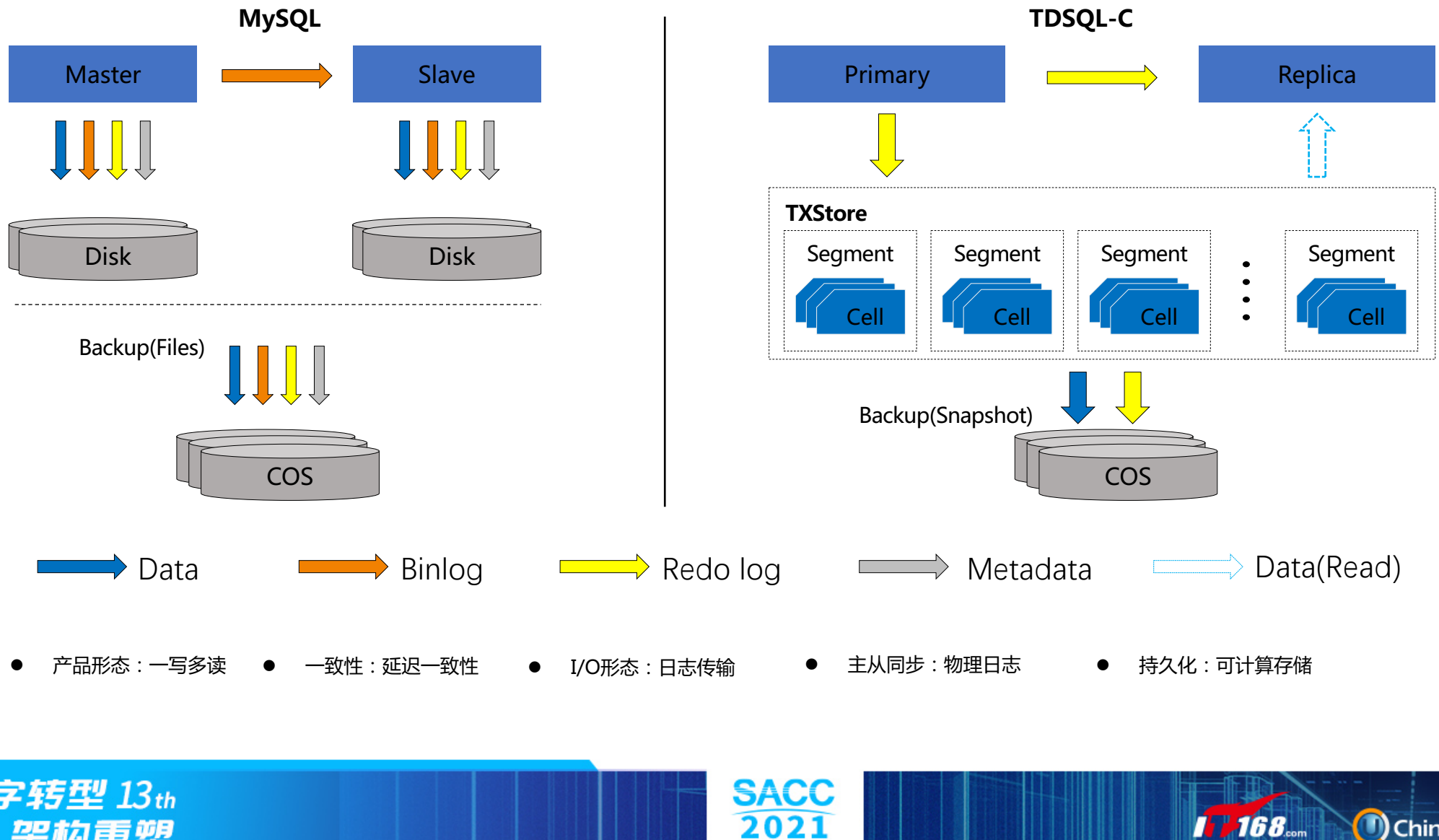
秒级 快照备份

Serverless

产品背景

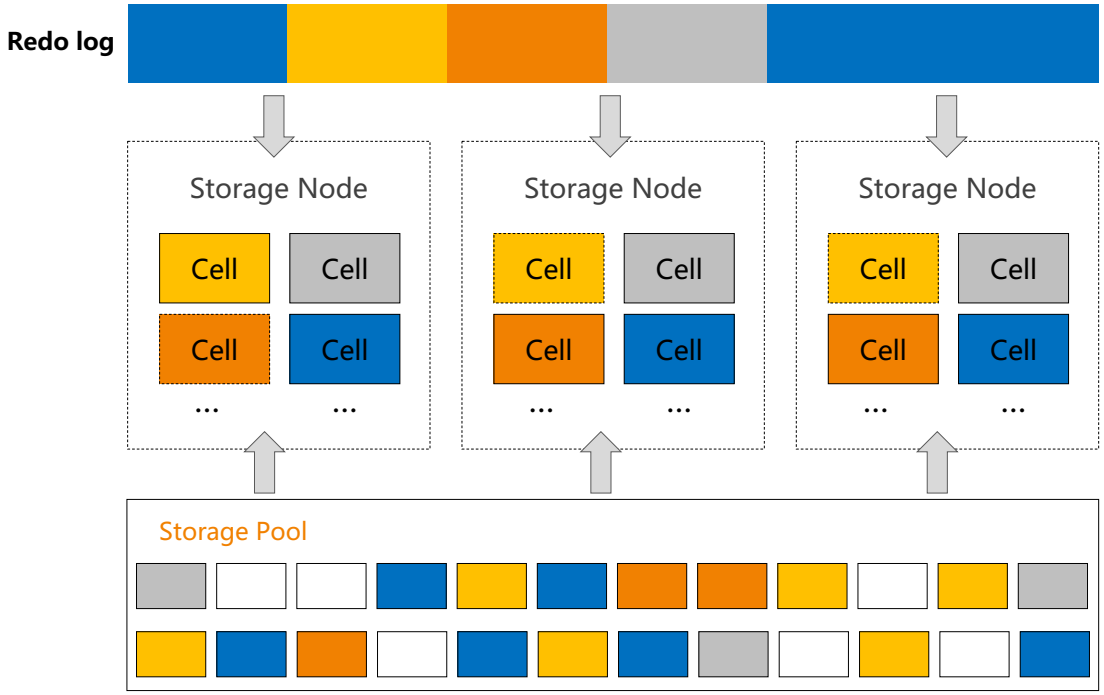
	存储容量	可靠性	可用性	水平扩展
传统架构问题	1、磁盘容量有限 2、扩容对业务影响大 3、分库分表对业务影响大，分布式事务问题多	1、普通复制（binlog）可能丢失数据（RPO>0） 2、同步复制性能差	1、HA、恢复速度慢（RTO分钟级） 2、副本时延大（分钟级-小时级）	1、水平扩展需要完整数据库副本，产生大量IO 2、只读副本部署速度慢（分钟级-小时级）
用户需求	1、大于100T容量 2、快速、透明扩容	1、不能丢失数据（RPO=0） 2、多副本容灾	1、快速HA、恢复（RTO秒级）、回档 2、更小的副本时延（毫秒级）	1、秒级副本扩展
技术方案	1、云存储：理论无上限，多副本可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

架构设计

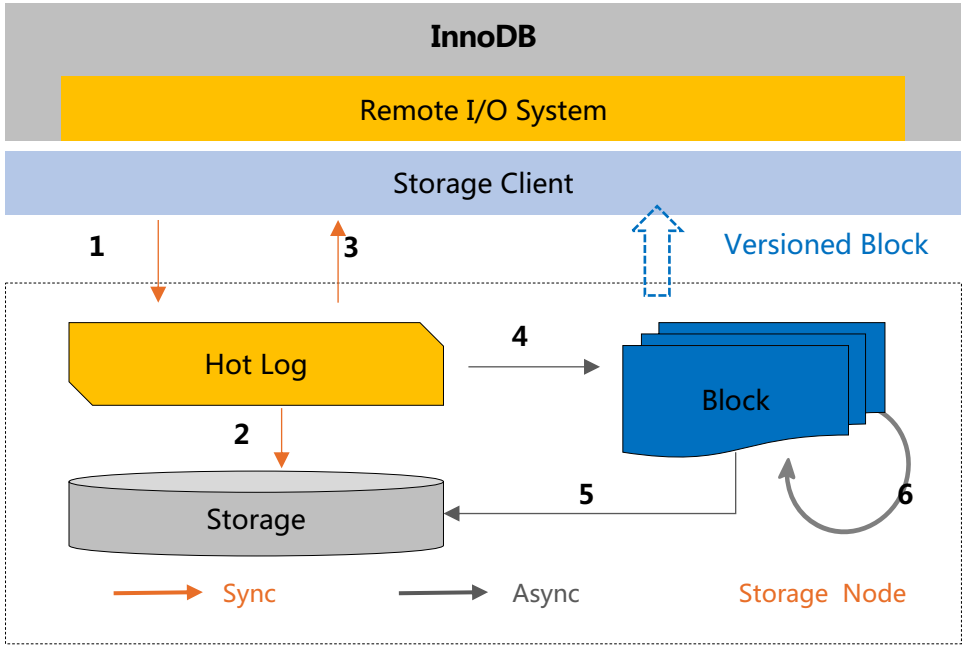


- 产品形态：一写多读
- 一致性：延迟一致性
- I/O形态：日志传输
- 主从同步：物理日志
- 持久化：可计算存储

可计算存储



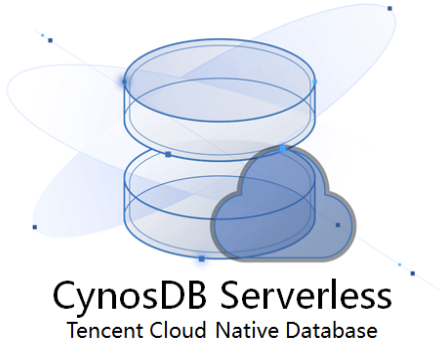
- ✓ 日志按照所属页面分片
- ✓ 分片包含独立的日志和数据
- ✓ 三副本存储
- ✓ 存储池最小1M物理分配单元



1. 传输日志到存储节点
2. 持久化日志
3. 通知客户端日志完成持久化
4. 回放日志到数据页面
5. 持久化新版本页面
6. 回收日志和页面

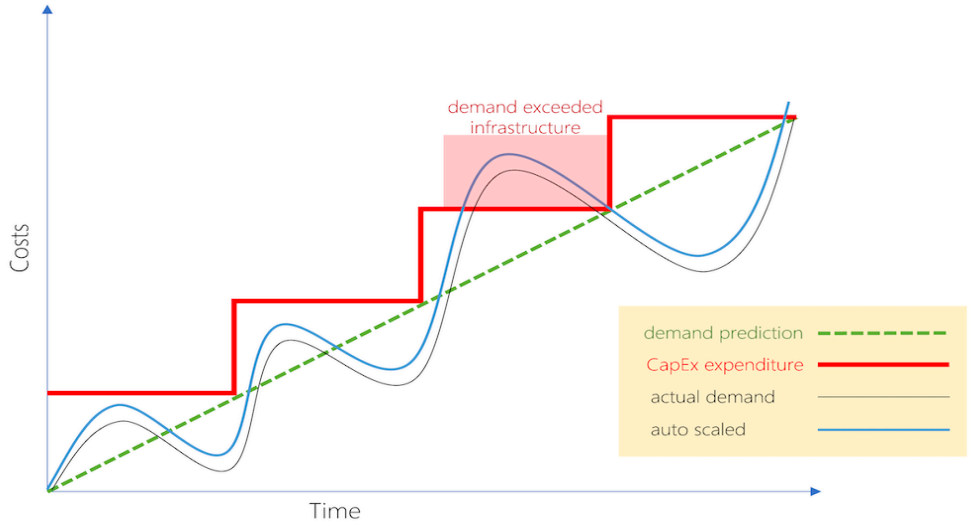
目录

突破一：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

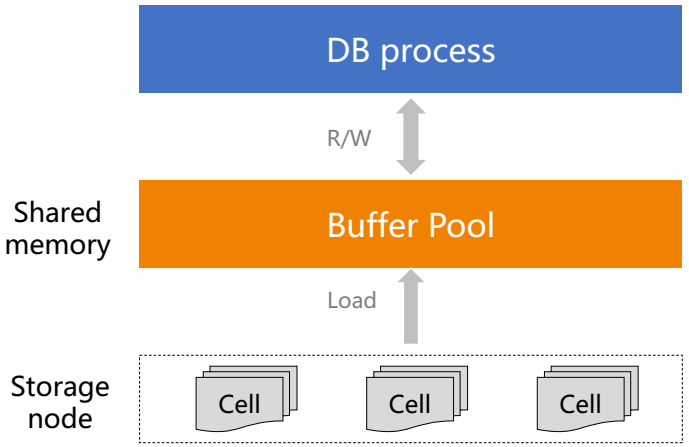


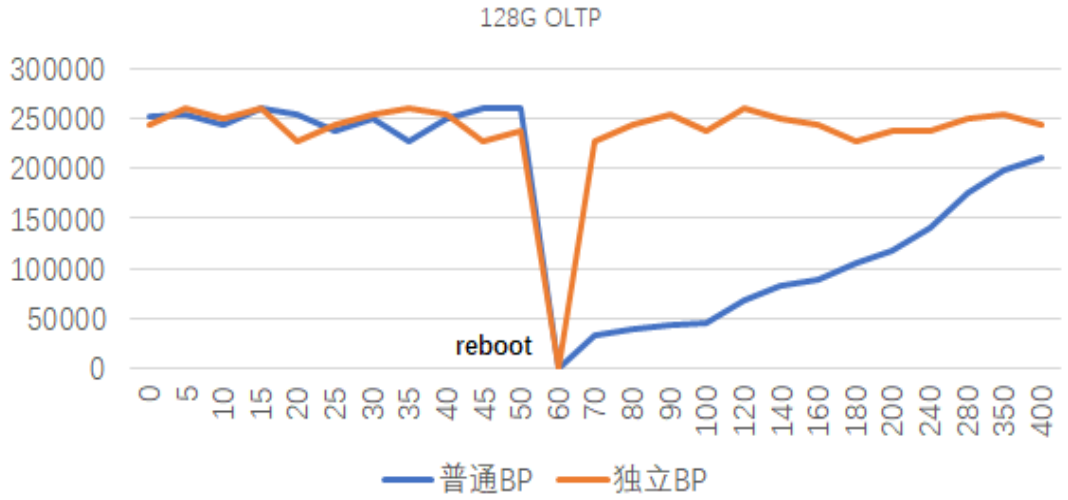
特性

智能极致弹性：极速启停，根据负载启停实例。无感知扩缩容，按需扩容，自动缩容

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。

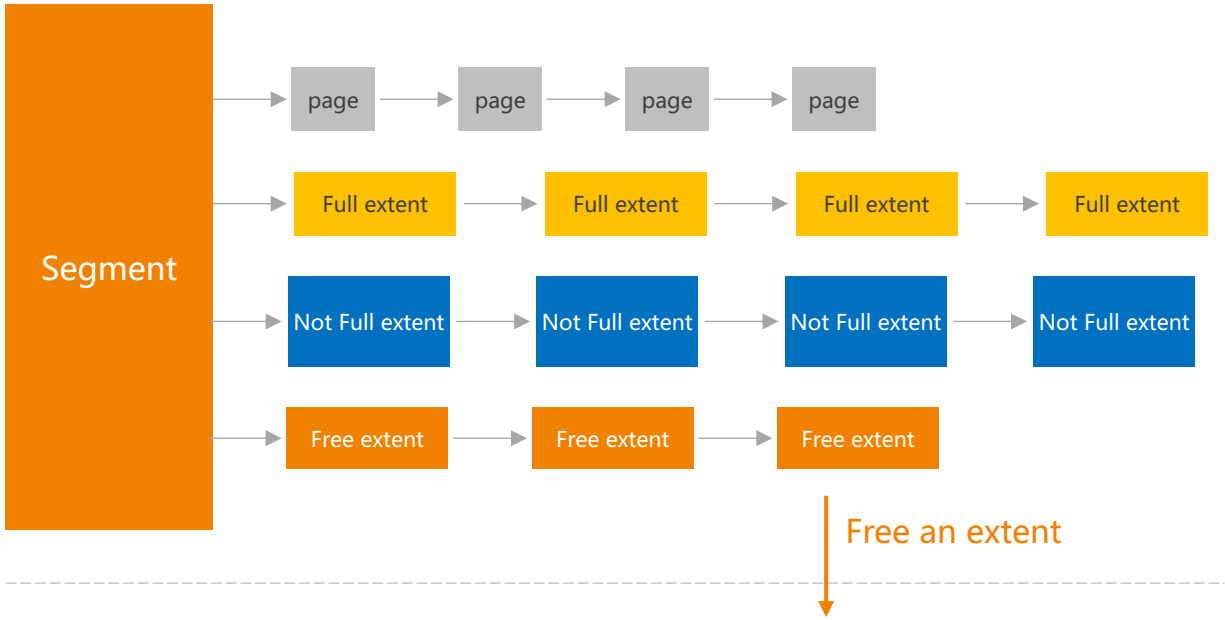
极速启停



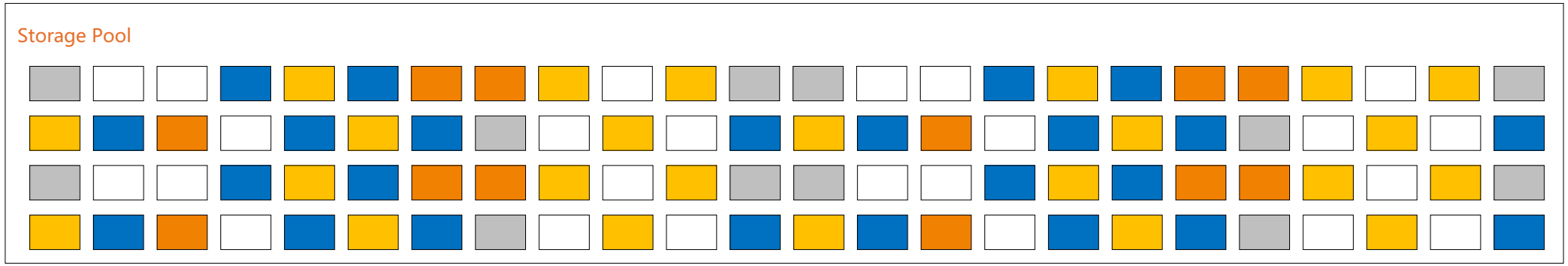


- 独立buffer pool
  - Buffer Pool使用共享内存，从计算节点分离
  - 缩短实例启动和恢复的时间，启动后性能无明显衰减
- 并行恢复
  - 计算层卸载恢复逻辑
  - 存储层多分片并行恢复
  - 页面版本化按需回放
- 启动优化
  - Buffer pool并行初始化
  - Rollback segment并行初始化
  - 表锁恢复优化
- 快速停机
  - 卸载刷脏，日志落盘即可停机

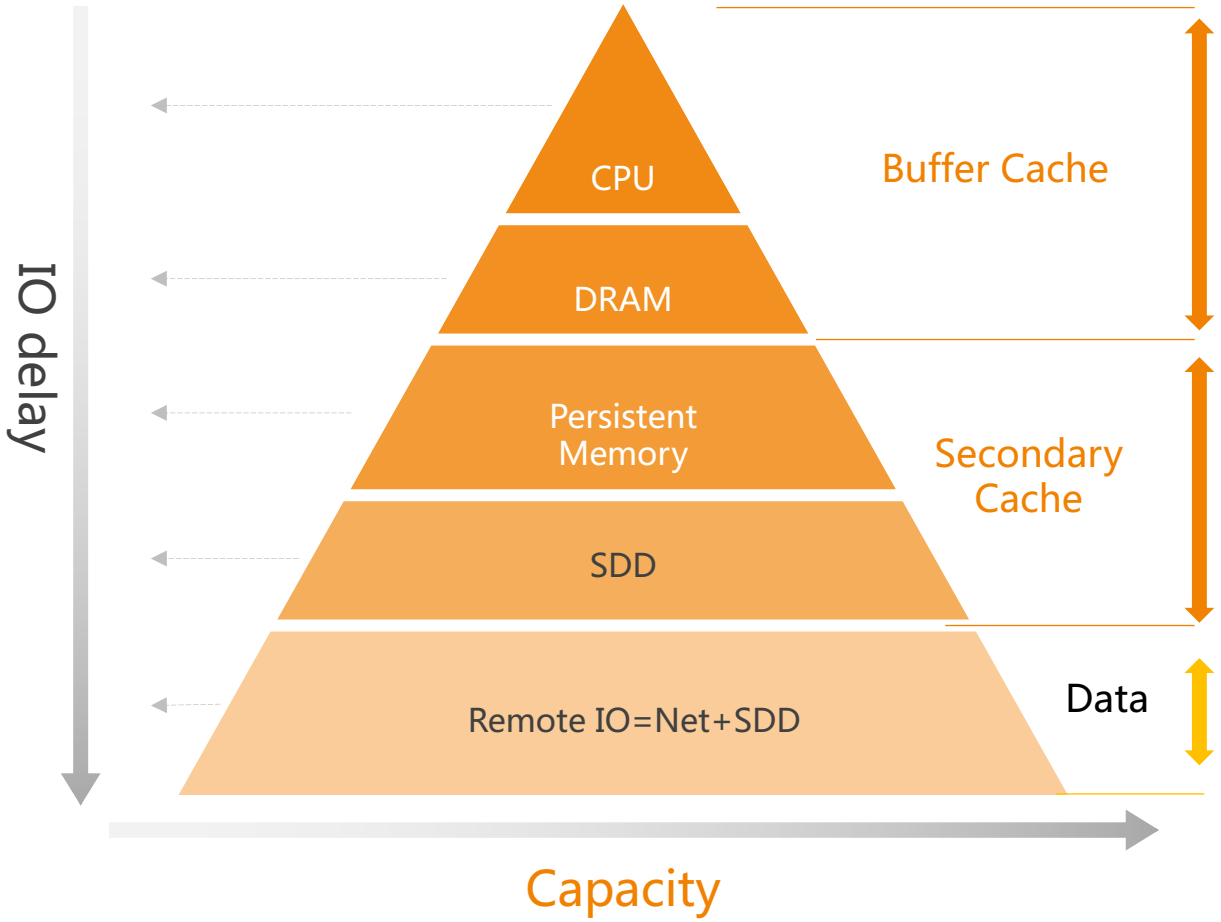
极致伸缩



- 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达到一定数量时触发存储池的回收
- 提供真正意义上的按需计费能力

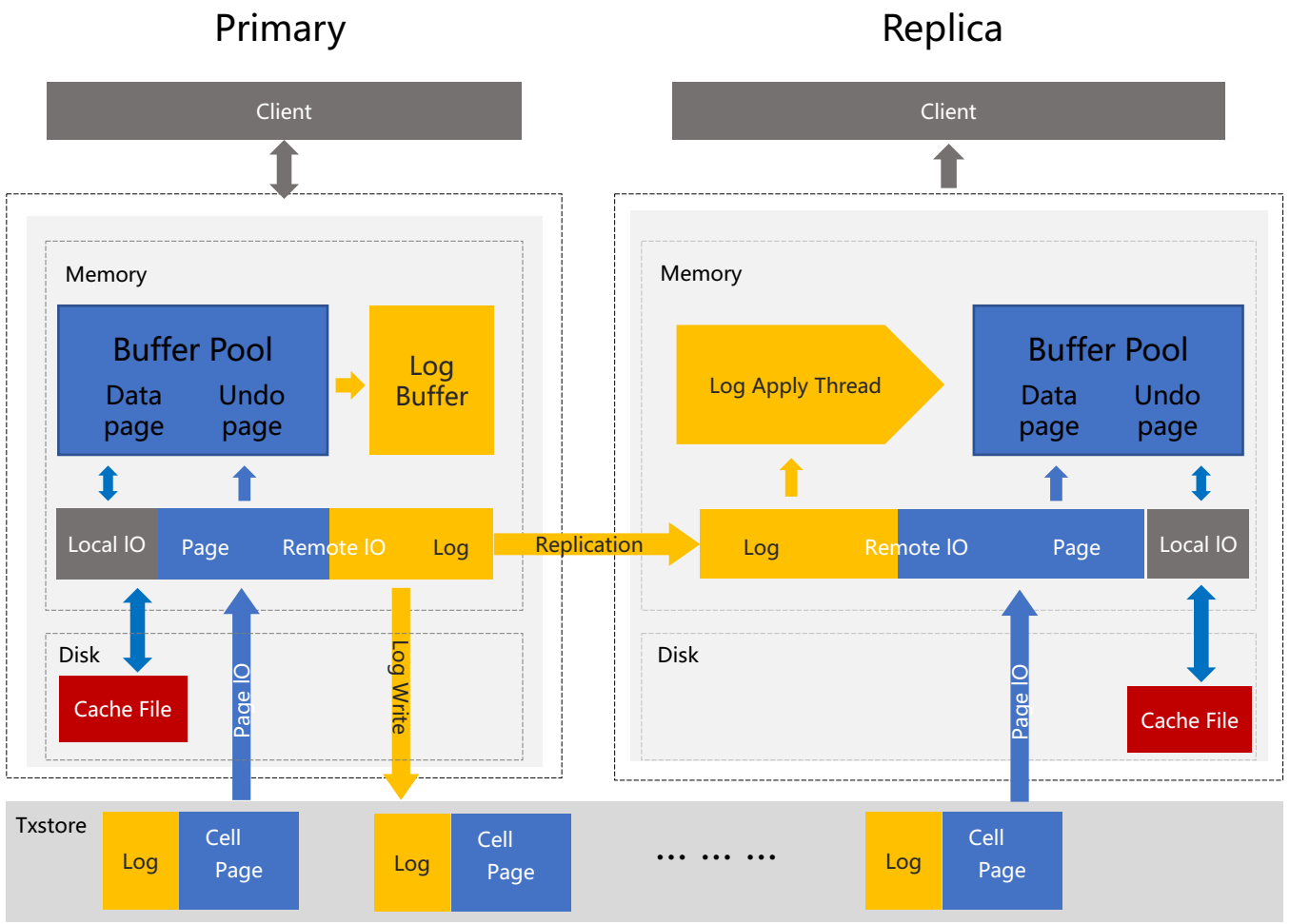


突破二：IO Bound优化



- 中小规格实例，内存规格远小于数据容量
- 超大规格实例，内存规格触及上线
- IO bound场景性能较差
- Remote IO放大IO bound场景影响

二级缓存



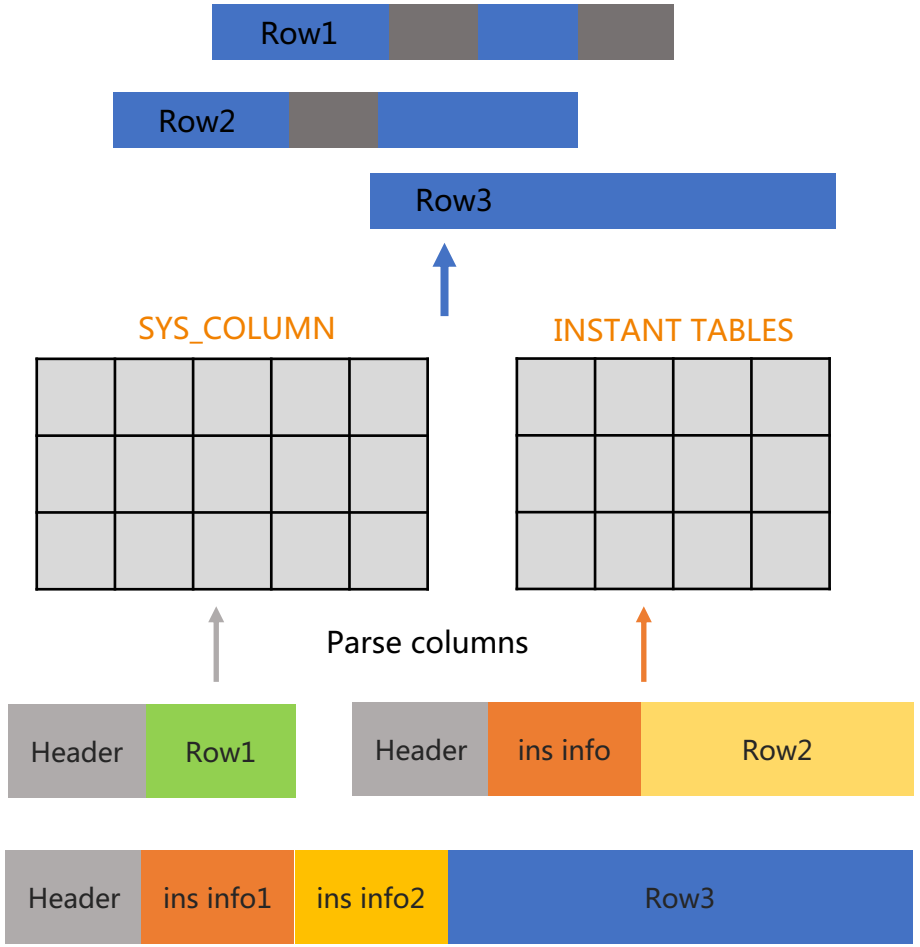
二级缓存：

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，提供快速高效的热数据访问能力

效果：

随着数据量的增大，性能平均提升100%以上

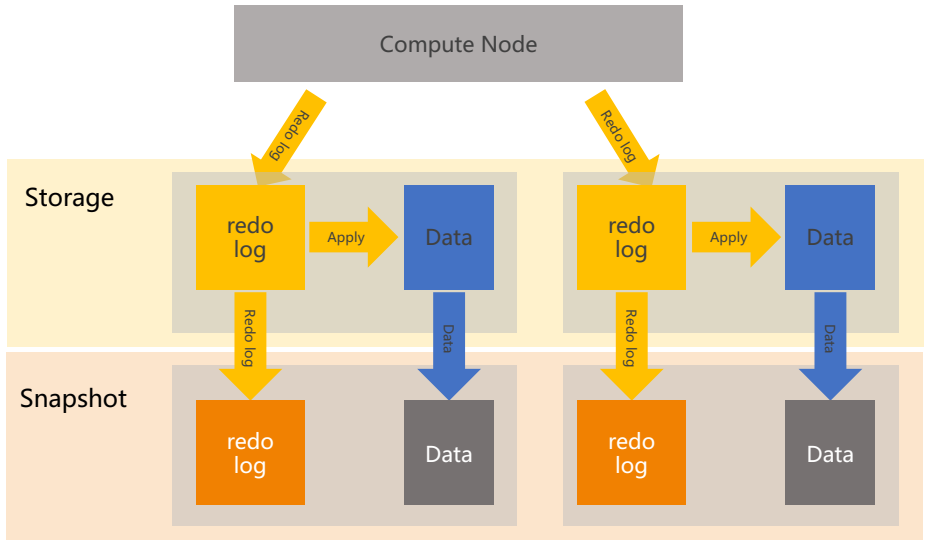
Instant DDL



- Instant DDL
  - 新增列
  - 修改列类型
  - 删除列
- 并行rebuild
  - 并行扫描
  - 并行导入

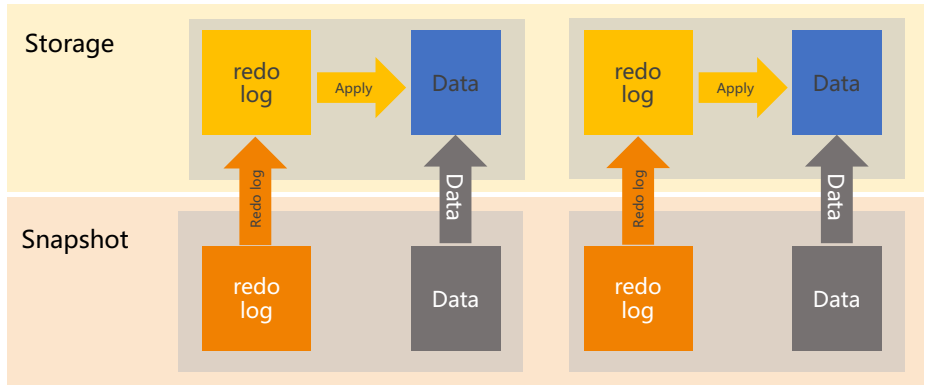
突破三：持续备份，并行回档

自治备份



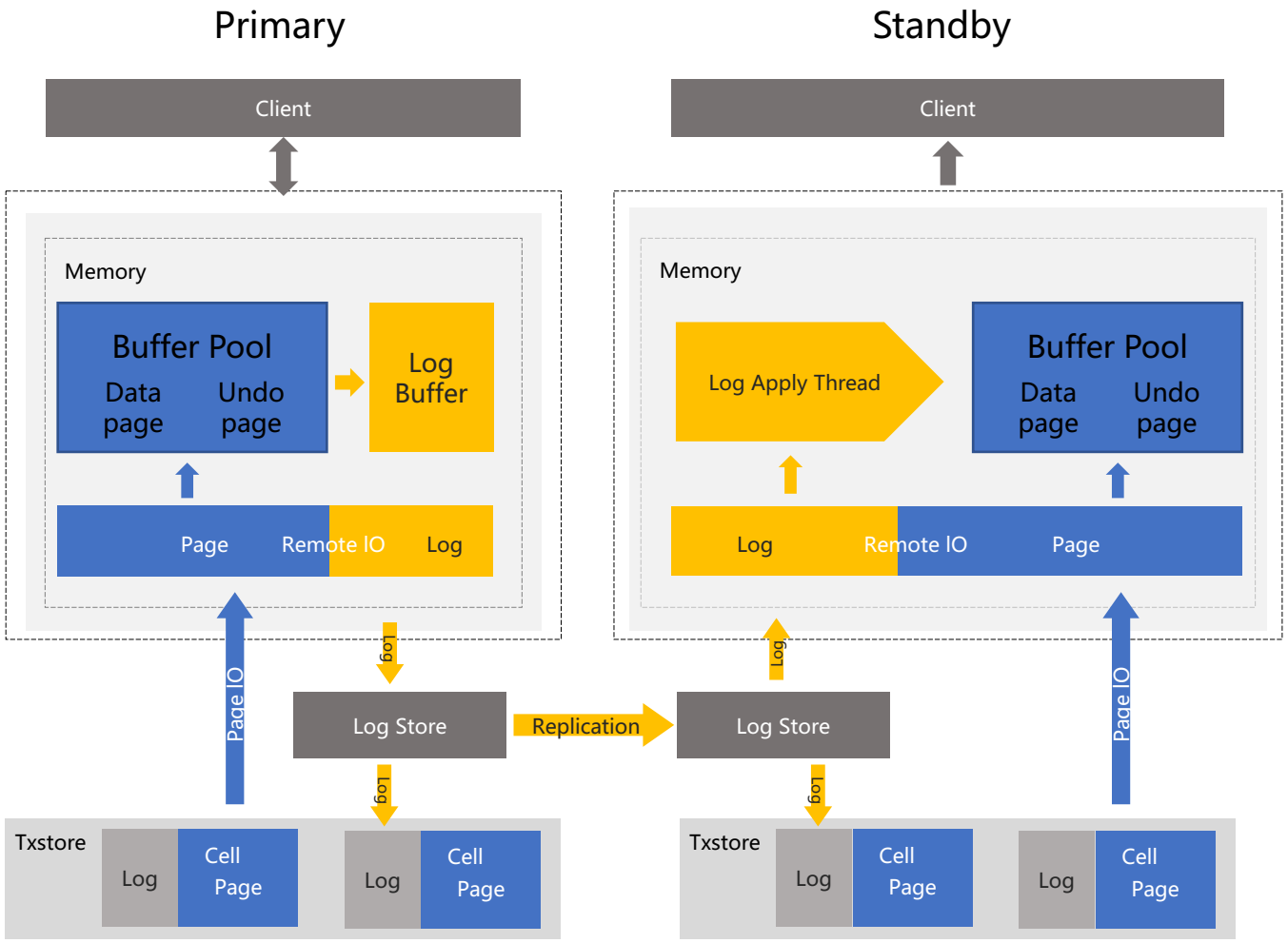
- 自治备份：存储分片根据备份点进行独立备份，同时做到备份全局一致性备份
- 并行回档：每个分片并行查找数据全量/增量备份，并行回放日志

并行回档



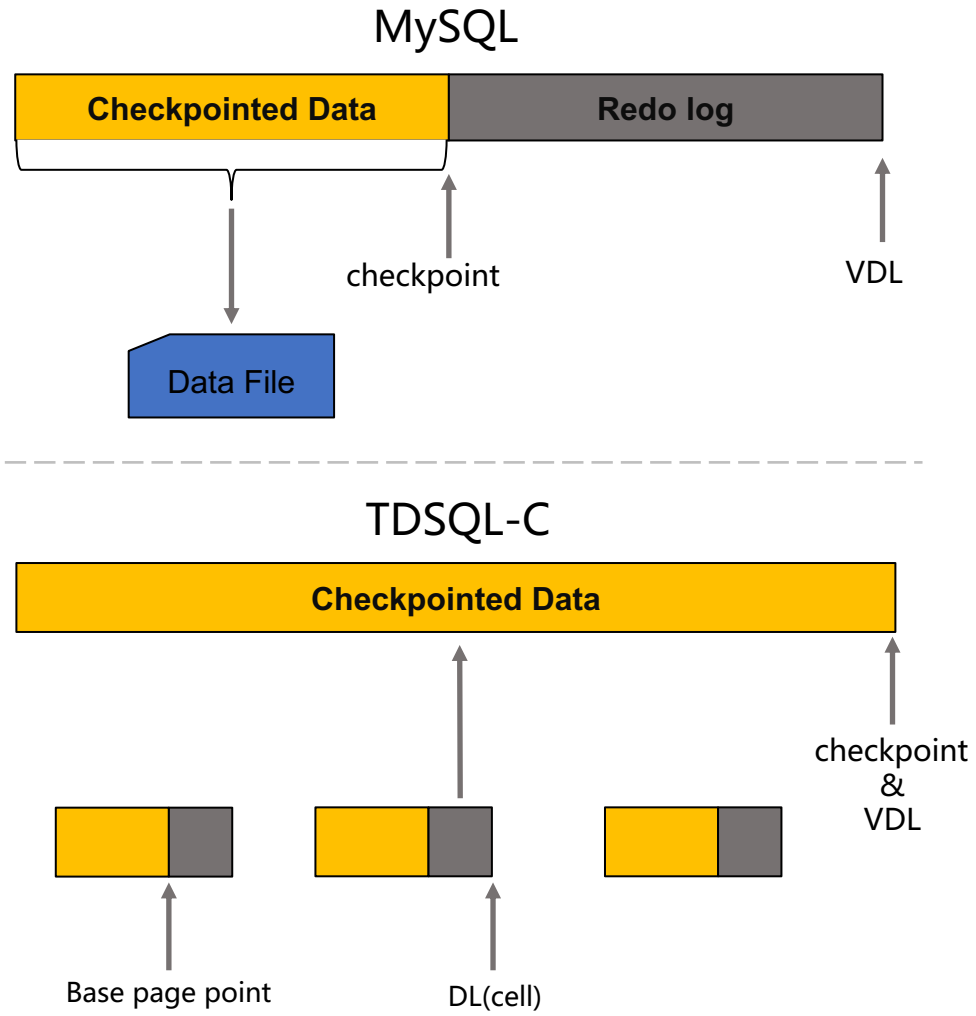
持续备份：持续的无感知备份，独立并行，秒级备份；并行回档，GB级回档速度

Global standby



- 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

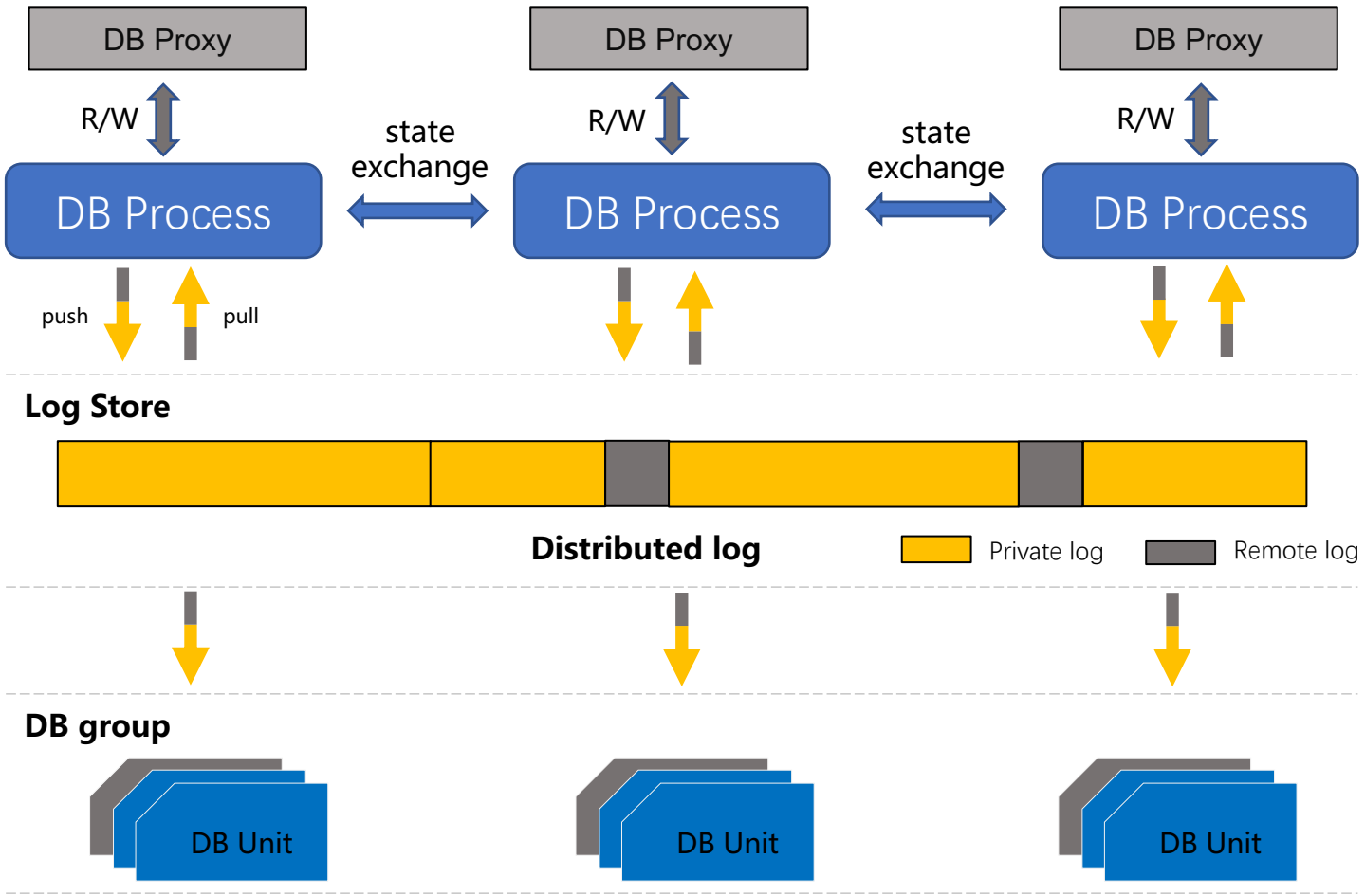
深度优化



打破常规：日志即数据和页面版本化，为内核的深度优化提供了新的方法

- 页面淘汰：页面淘汰不再与日志持久化关联，快速淘汰页面，保证并发稳定性
- 分区读写：多线程对一个页面进行分区读写，提升读写并发能力
- 远程写：特定页面远程写

分区多写



- 数据集分区
- 多节点读写
- 日志传输
- 全局事务

关注“**腾讯云数据库**”官方微信

体验**小程序一键管理数据库**

获取数据库技术干货和最新资讯


SACC
THANKS