



PostgreSQL中文社区



PostgreSQL中文社区

2021 PostgreSQL China Conference
主办：PostgreSQL 中文社区

第11届 PostgreSQL 中国技术大会

开源论道 × 数据驱动 × 共建数字化未来





2021 PostgreSQL China Conference
第 11 届 PostgreSQL 中国技术大会



PostgreSQL 中文社区

TDSQL-C for PG 主从架构优化

演讲人：邹立贤 腾讯云



1. TDSQL-C for PG 整体架构



2. TDSQL-C for PG 主从架构



3. TDSQL-C for PG 优化



4. 展望



1. TDSQL-C for PG 整体架构



为什么需要 TDSQL-C

传统数据库在云上存在问题

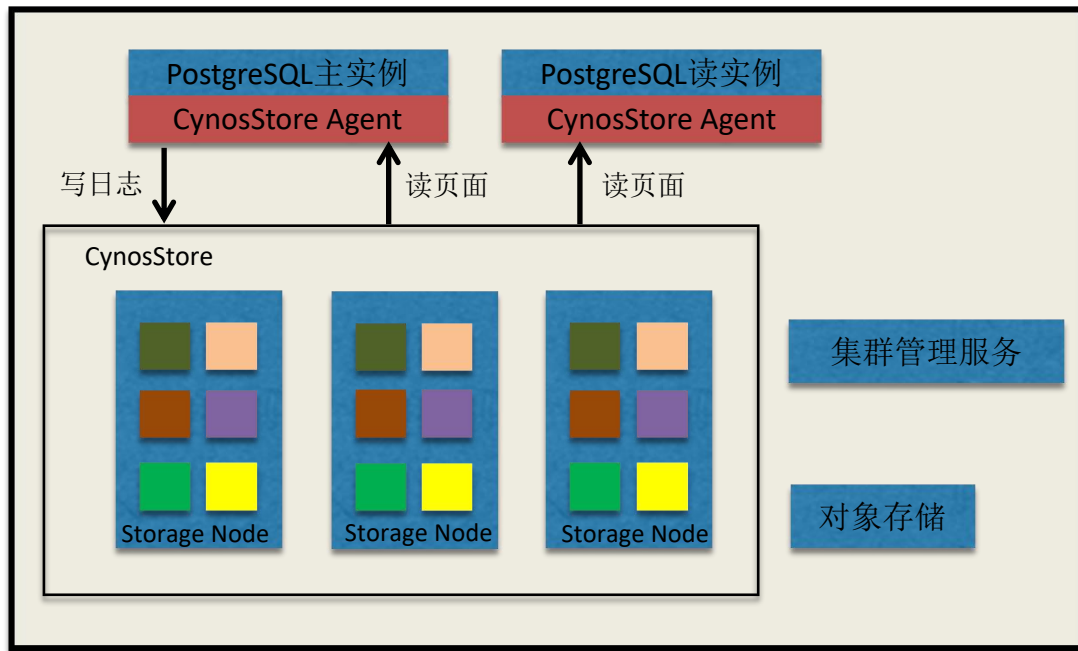
- 资源利用率低
- 扩展能力不足
- 资源规划难
- 备份难

TDSQL-C 解决思路

- 计算存储分离
计算资源弹性调度能力
- 日志下沉以及异步回放
减少网络IO
- 共享分布式存储
资源弹性扩展
- 后台持续日志备份



TDSQL-C for PG 架构

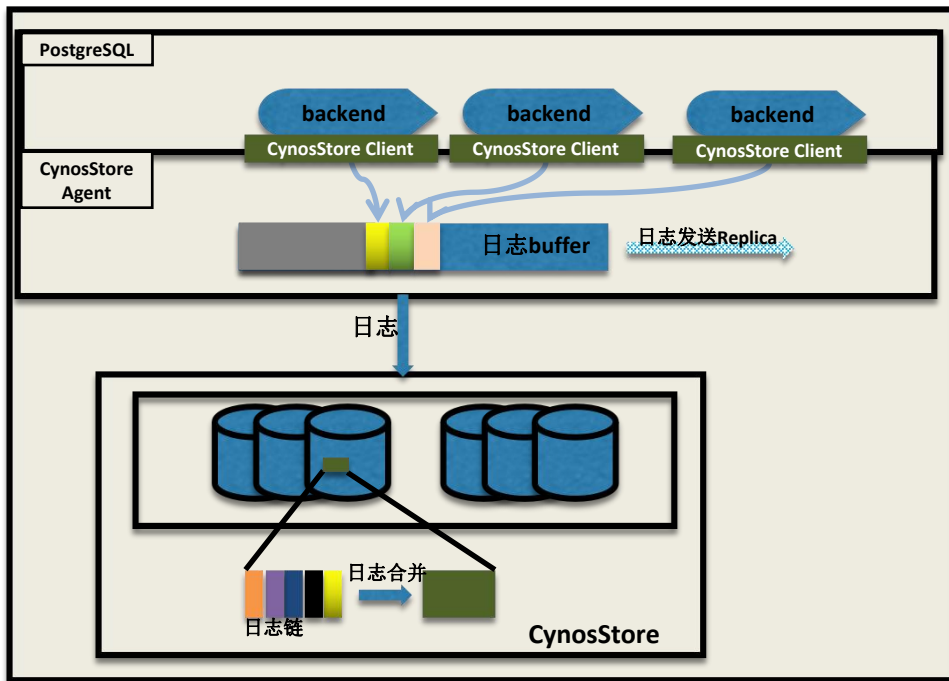


- 核心架构设计
 - 日志下沉
 - 日志异步回放
 - 多版本读 (同步)



TDSQL-C for PG 架构

- 日志下沉、异步回放



- 数据原子修改 (MTR)
 - MTR (Minimal Transaction Record)
 - CPL (Consistency Point LSN)
 - VDL (Volume Durable LSN)
- 日志异步写入
 - 日志写入buffer, 再写入存储
- 日志并行插入



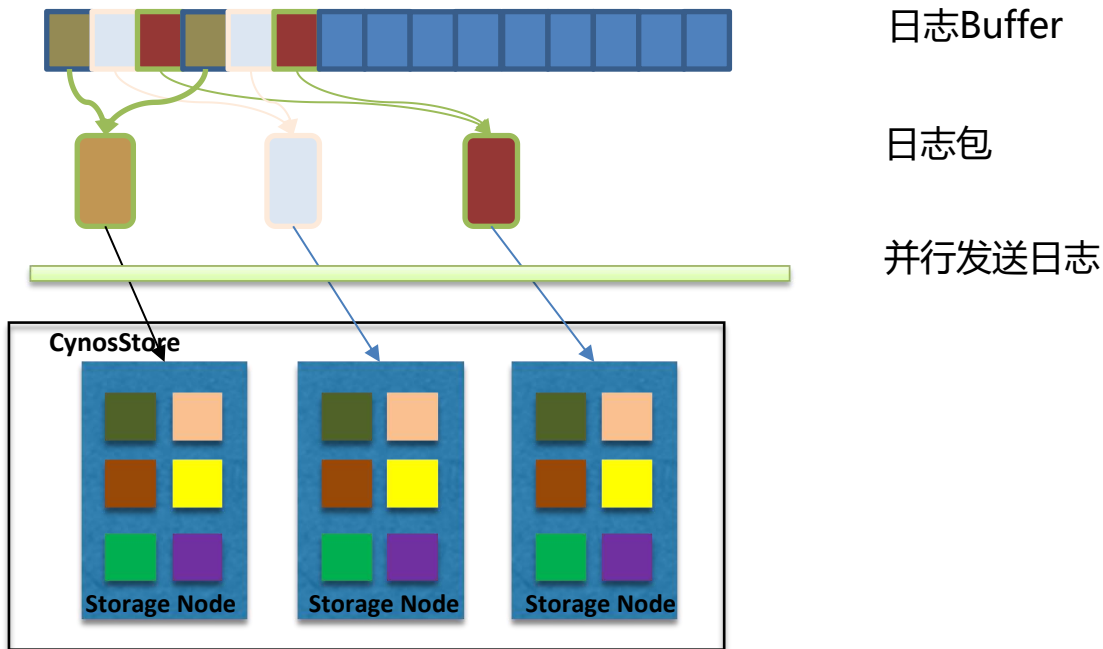
TDSQL-C for PG 主机优化

- ✓ 不必依赖 CheckPoint 机制
- ✓ 不必在存储日志中记录全页(FPW)
- ✓ 快速启动系统，无需恢复大量日志
- ✓ 批量并行发送日志到存储
- ✓ 日志合并压缩，进一步减少日志IO



TDSQL-C for PG 主机优化

批量并行发送日志到存储



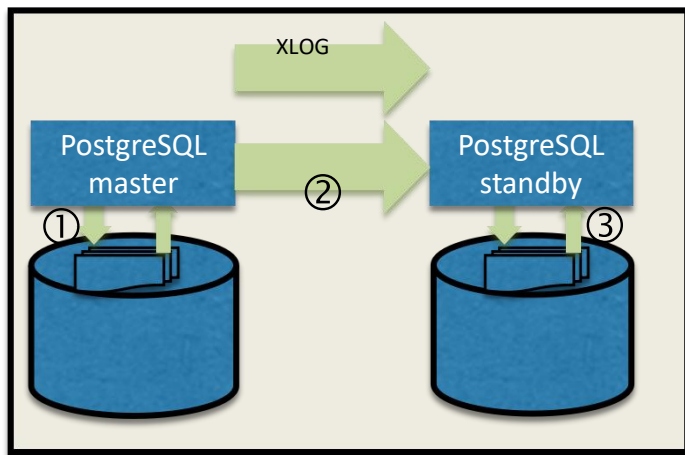


2. TDSQL-C for PG 主从架构



TDSQL-C for PG 主从架构

- 传统PostgreSQL主备模式仍有缺点

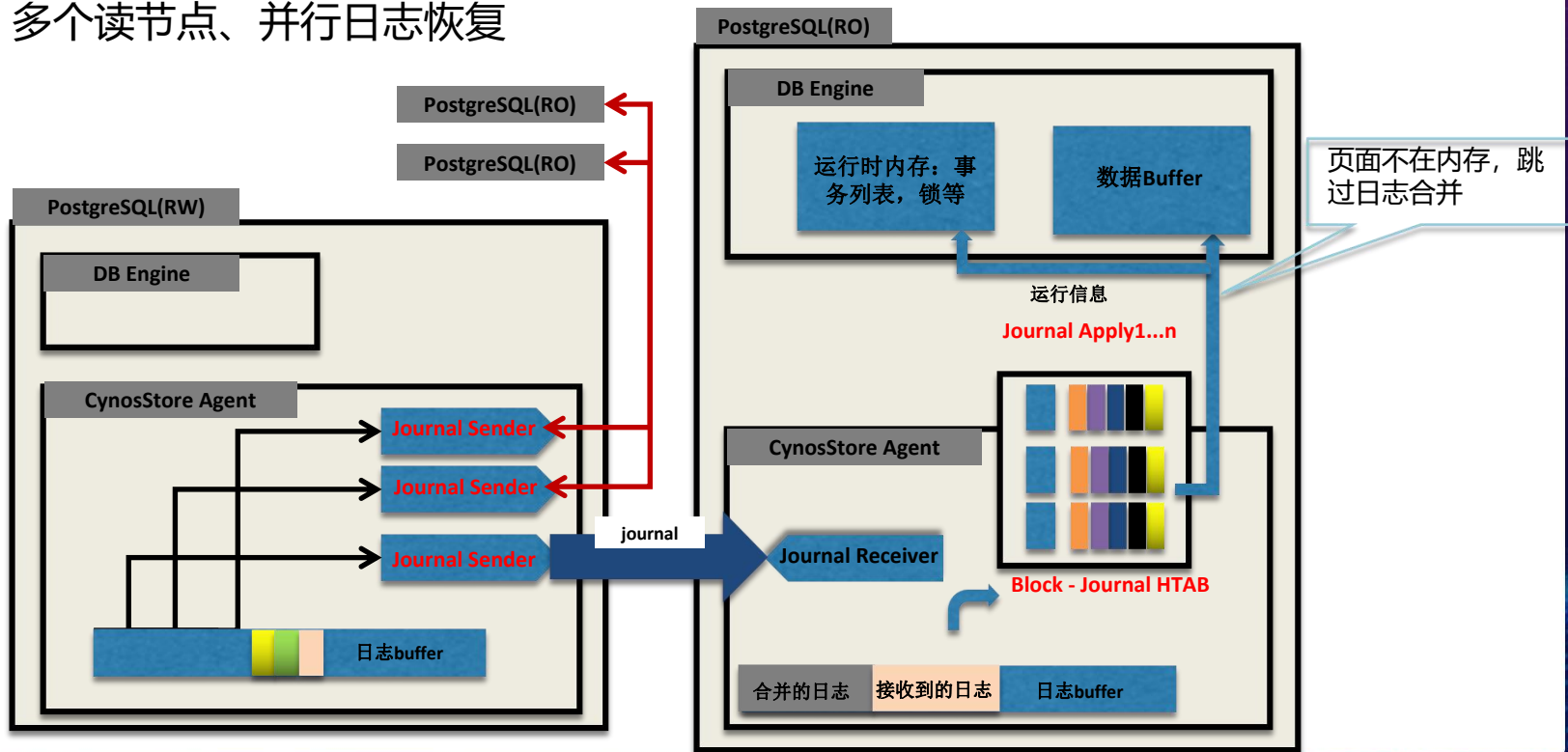


- 传统主备模式的问题
 - 创建新备需要拷贝数据：额外存储资源
 - 备机切换和启动需要恢复日志：慢
 - 冲突导致的日志延迟等
- 多读 (Replica) 优势
 - 横向扩展读能力
 - 提升数据库的可用性



TDSQL-C for PG 主从架构

多个读节点、并行日志恢复



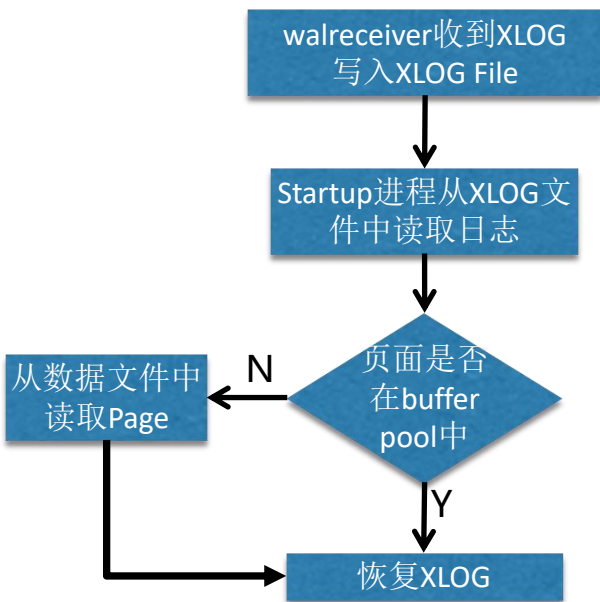


3. TDSQL-C for PG 优化

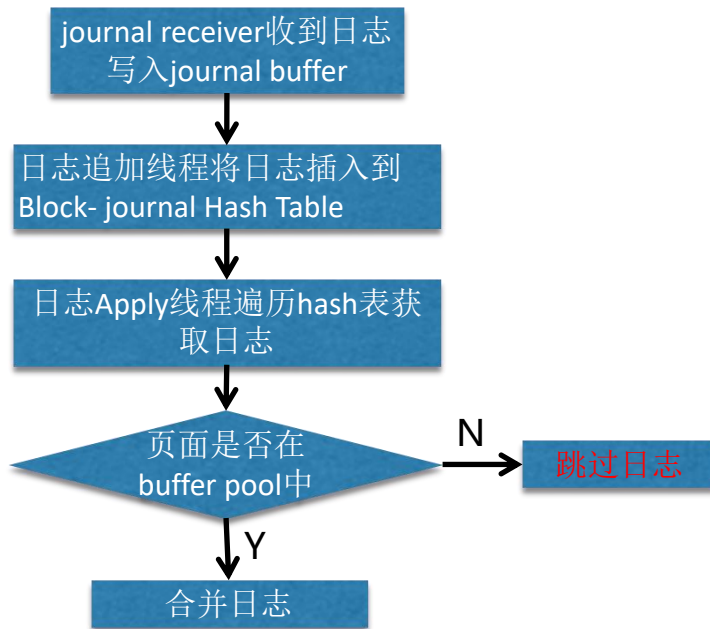


TDSQL-C for PG 从机优化

传统PostgreSQL恢复过程



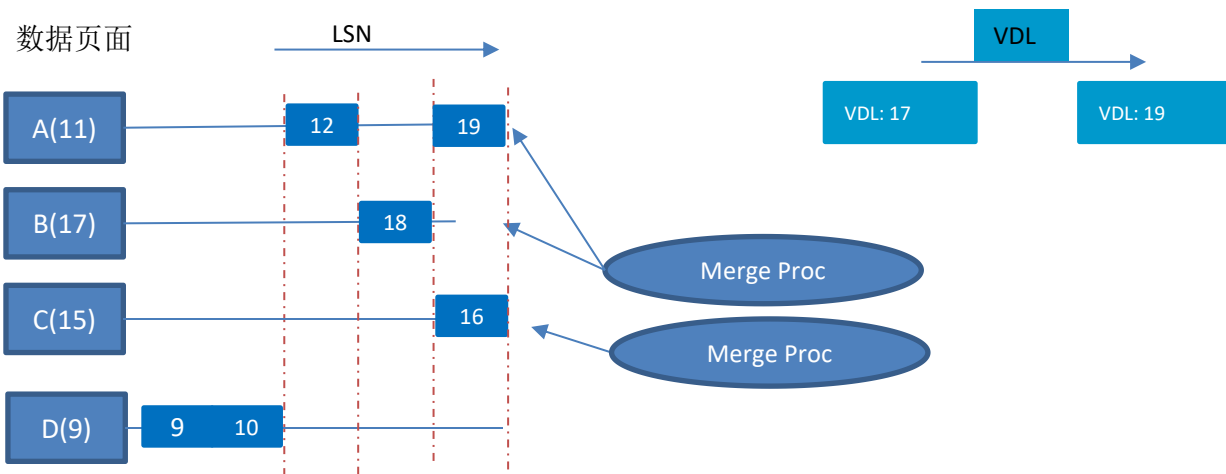
TDSQL-C for PG





TDSQL-C for PG 从机优化

并行合并日志





TDSQL-C for PG 从机优化

PostgreSQL处理DROP表/数据库流程:

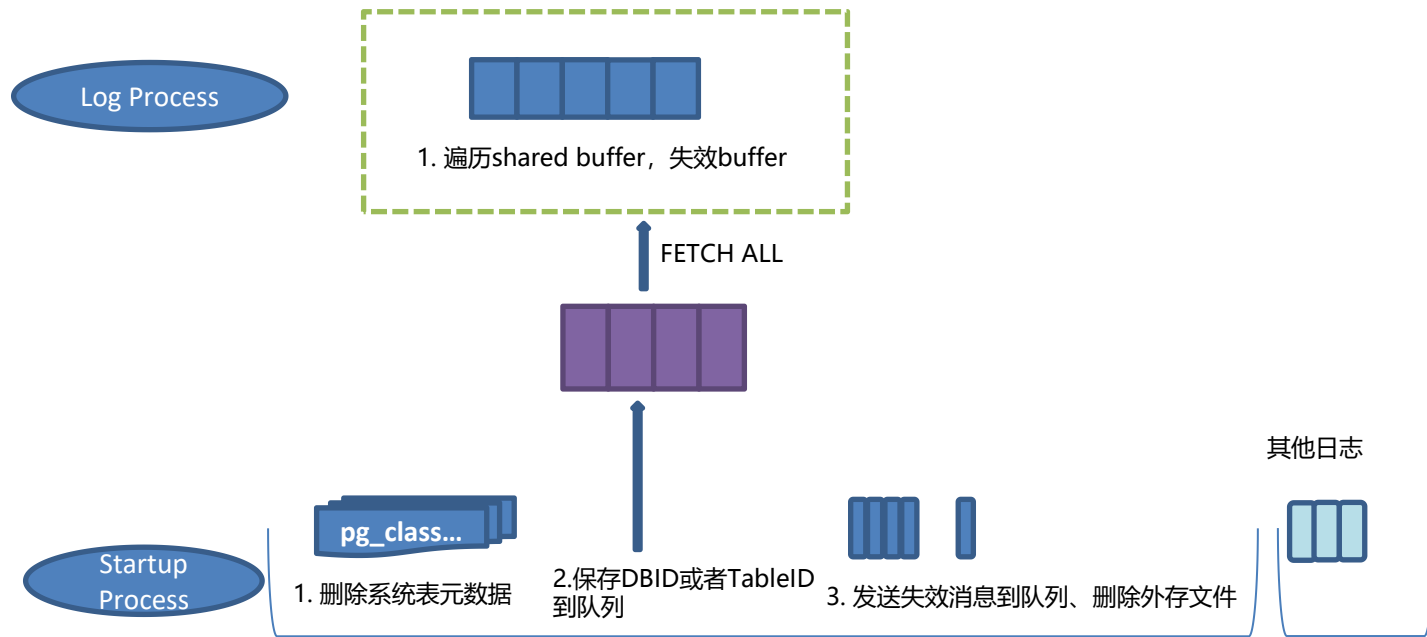
- ❑ XLog逐条在Startup中处理
- ❑ Startup进程是CPU瓶颈





TDSQL-C for PG 从机优化

TDSQL-C处理DROP表/数据库优化思路





TDSQL-C for PG 从机优化

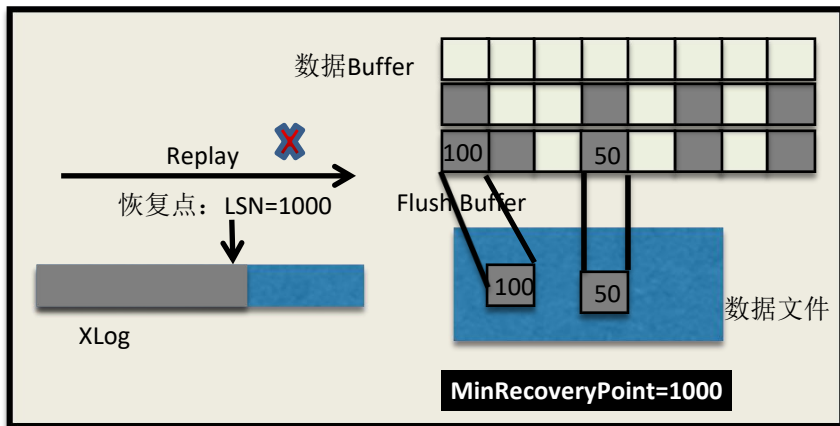
- PostgreSQL 从机删除表在内存中的信息
 - 原本是一个单向列表保存表相关信息
 - 从列表中顺序遍历找到要关闭的表
 - 找到后删除
- TDSQL-C优化后
 - 使用双向列表，加快删除效率
- TDSQL-C优化后效果，日志堆积从100GB级别基本降低为MB级别



TDSQL-C for PG从机优化

- 传统PostgreSQL备机的启动

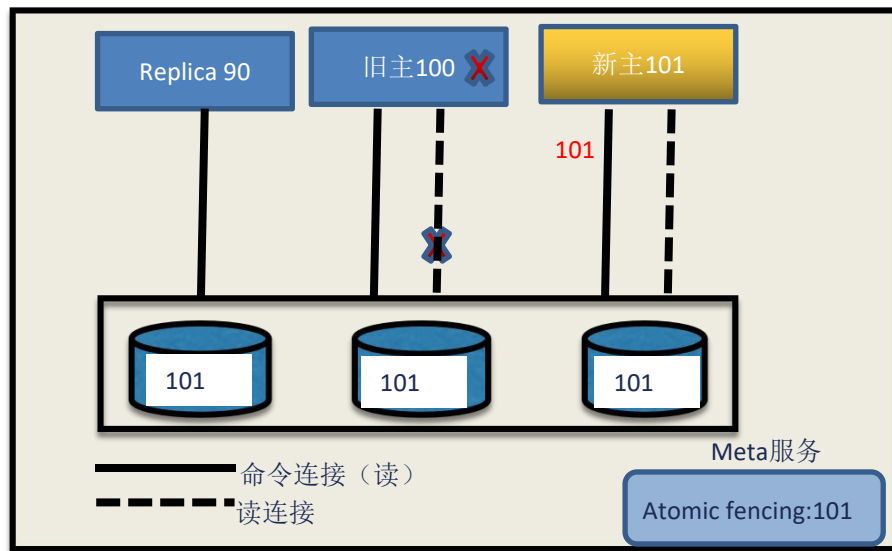
- 传统备机启动需要恢复到一致点才能保证数据正确并提供读服务



- TDSQL-C for PG 从机启动需要一个持久化的VDL就可以获取存储的一致状态
- 快速创建Replica从机，无需复制全量数据



TDSQL-C for PG主从优化



- 避免传统主备模式下，出现双写导致日志“分叉”
- HA Fencing机制
 - 数据库启动，获取当前fencing 值
 - 使用获取的fencing值连接到存储



TDSQL-C for PG 更多探索

1

AEP & Optane :

提升存储性能, 降低延迟

2

40G/100G+RDMA :

提升网络性能, 降低延迟

3

Multi-Master & Serverless:

多主架构, 无服务化, 进一步降低计算成本

4

易用性:

Oracle兼容



2021 PostgreSQL China Conference
第 11 届 PostgreSQL 中国技术大会



PostgreSQL 中文社区

THANKS

开源论道 × 数据驱动 × 共建数字化未来