

TiDB DR–Auto–Sync 同城双中心的原理与实践

TiDB Club 2024年11月15日 17:05 北京



微信扫一扫
关注该公众号



👉 点击此处，立即报名 👉



本合集精选两篇深度文章，介绍了 DR Auto–Sync 技术如何保障同城双中心高可用性，探索 TiDB 在分布式数据库领域的创新实践，希望大家提供 TiDB 技术原理的全面视角，助力优化数据库运维管理。

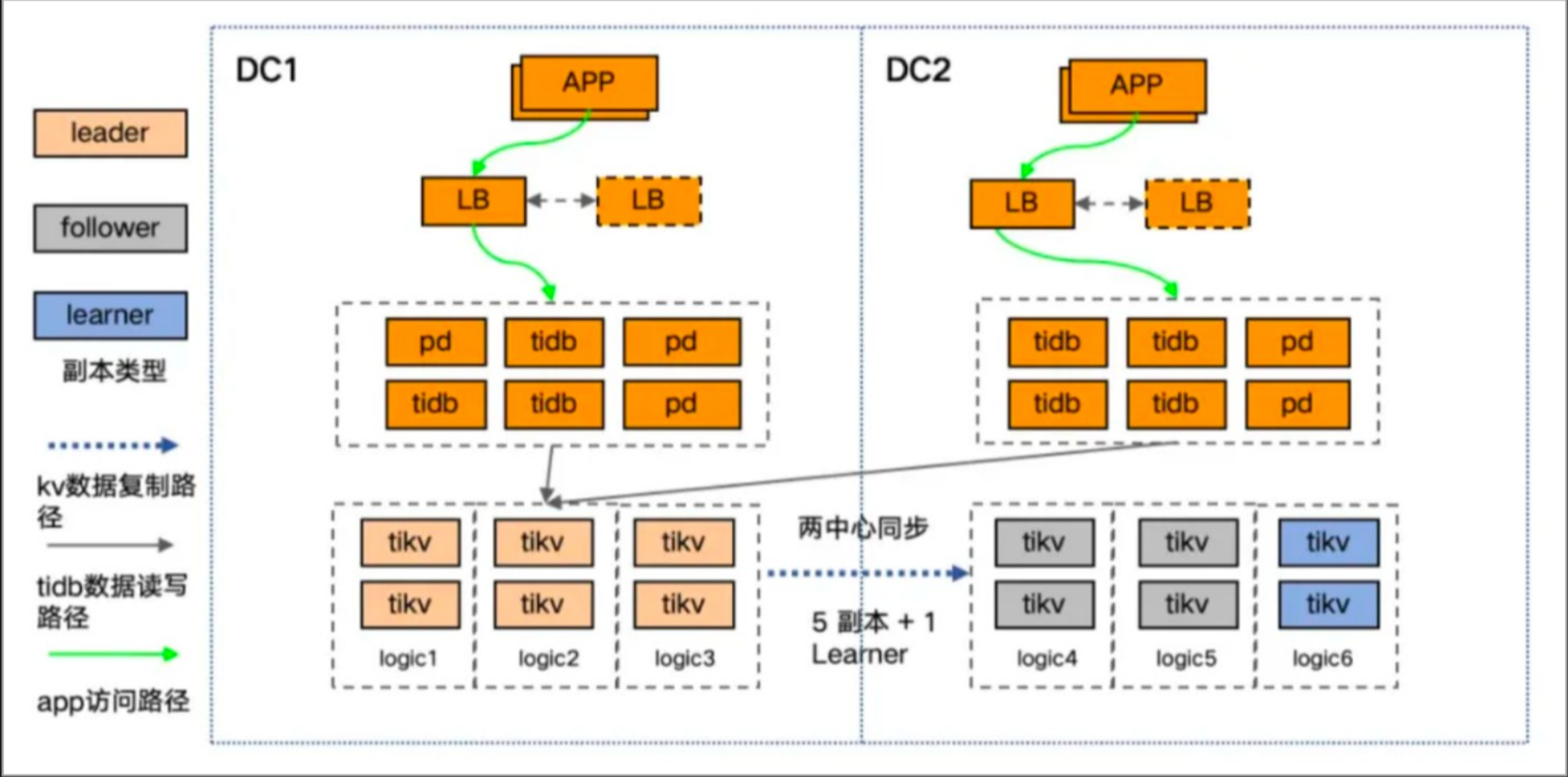
01

TiDB DR–Auto–Sync 同城双中心高可用实践

银行核心背后的落地工程体系

DR–Auto–Sync 简介和部署架构

DR–Auto–Sync 适用于同城双中心，需 50 公里内、低时延高带宽。支持状态自动切换，6 副本、3 Voter 主中心、2 Follower、1 Learner 副中心，PD 5 实例主中心，TiDB 双活。



集群配置


```
server_configs:
  tidb:
    proxy-protocol.networks: 192.168.1.70    #IP透传能力，硬件负载均衡
  tikv:
    raftstore.pd-store-heartbeat-tick-interval: 2s
    raftstore.snap-generator-pool-size: 1    #调整这个和下面的参数降低
    server.snap-max-write-bytes-per-sec: 30MB
    server.concurrent-send-snap-limit: 4
    server.concurrent-recv-snap-limit: 4
  pd:
    replication.max-replicas: 5    ##副本数量
    replication.location-labels: ["dc", "logic", "host"]
    replication.isolation-level: "logic"
tikv_servers:
- host: 192.168.1.1
  config:
    server.labels:
      dc: dc1
      host: "192_168_1_1"
      logic: logic1
  arch: arm64
  os: linux
```

集群拓扑文件

配置双中心 6 副本集群时，需调整 TiKV 心跳频率、snapshot 参数以避免自动切
换问题，并为 TiKV 实例设置数据中心和逻辑单元标签。

More 

Dr-Auto-Sync 容灾切换

DR-Auto-Sync 集群支持计划内切换和计划外切换（故障抢修）两种模式的容灾切换操作。

- **计划内切换和实战演练**：有计划的主中心、副中心角色转换，常用于日常容灾演练、副中心业务验证、计划内机房硬件维护等场景，操作简单、业务无感知。

- **计划外切换和实战演练**：主中心故障时，人工快速恢复 TiDB 集群，实现 RPO = 0、RTO < 1 min（从中心故障） 或 RTO < 10 min （主中心故障）的容灾保障。



● 切换方式实战效果对比

	日常运行	计划内切换		计划外切换	
		应用操作	RTO	应用操作	RTO
DR-Auto-Sync	双中心读写、业务流量可以跨中心调拨	无需操作	≈1 秒 交易无感知	重启切换至从中心数据源	分钟级别
主从方式的数据库	主中心读写、从中心只读业务	数据库角色和DNS 切换后, 应用重启切换数据源	分钟级别	重启切换至从中心数据源	分钟级别

◉ 监控与告警 ◉

TiDB v6.5.6 起，Grafana 集成 DR–Auto–Sync 监控，展示集群同步状态。监控重点包括主副中心数据同步、Region Leader 和 PD Leader 分布，设置告警：集群状态异常、副中心 Leader 存在时触发告警，PD Leader 状态变化用于切换判断，确保运维及时响应。

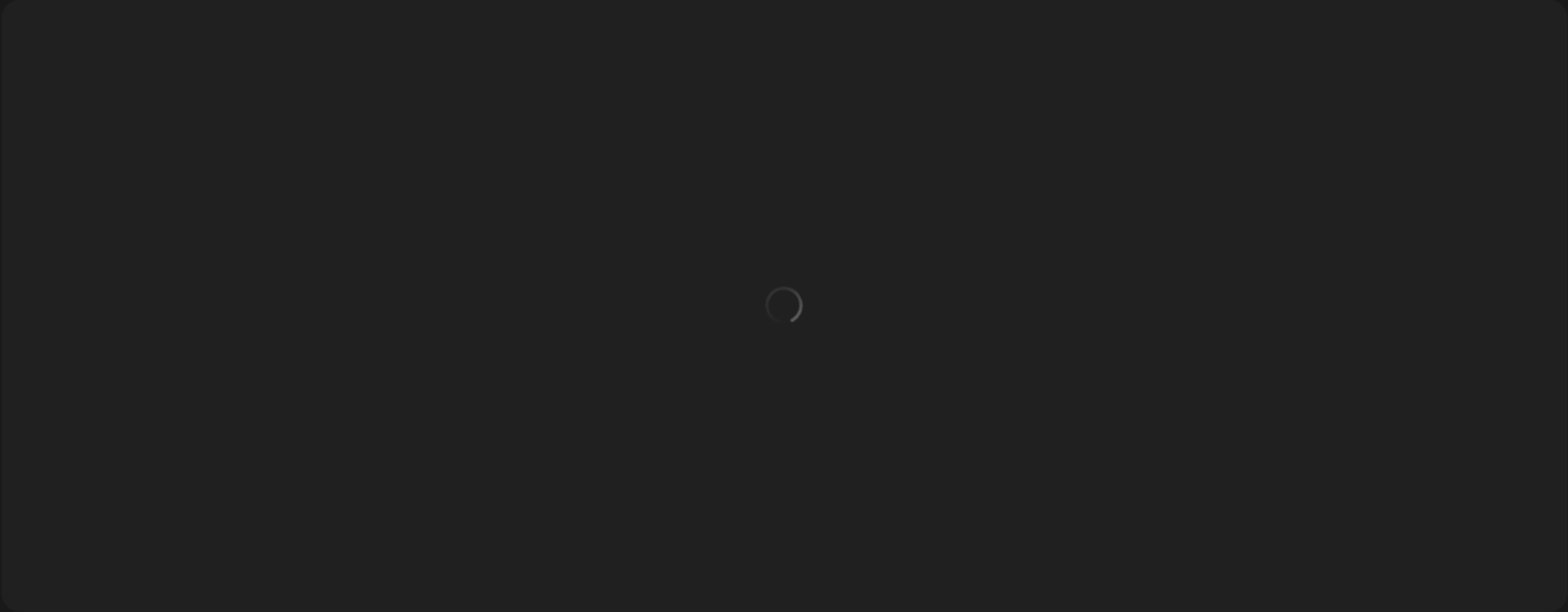


◉ 注意事项 ◉

DR-Auto-Sync	双中心读写、业务流量可以跨中心调拨	无需操作	≈1 秒 交易无感知	重启切换至从中心数据源	分钟级别
主从方式的数据库	主中心读写、从中心只读业务	数据库角色和 DNS 切换后，应用重启切换数据源	分钟级别	重启切换至从中心数据源	分钟级别

◉ 监控与告警 ◉

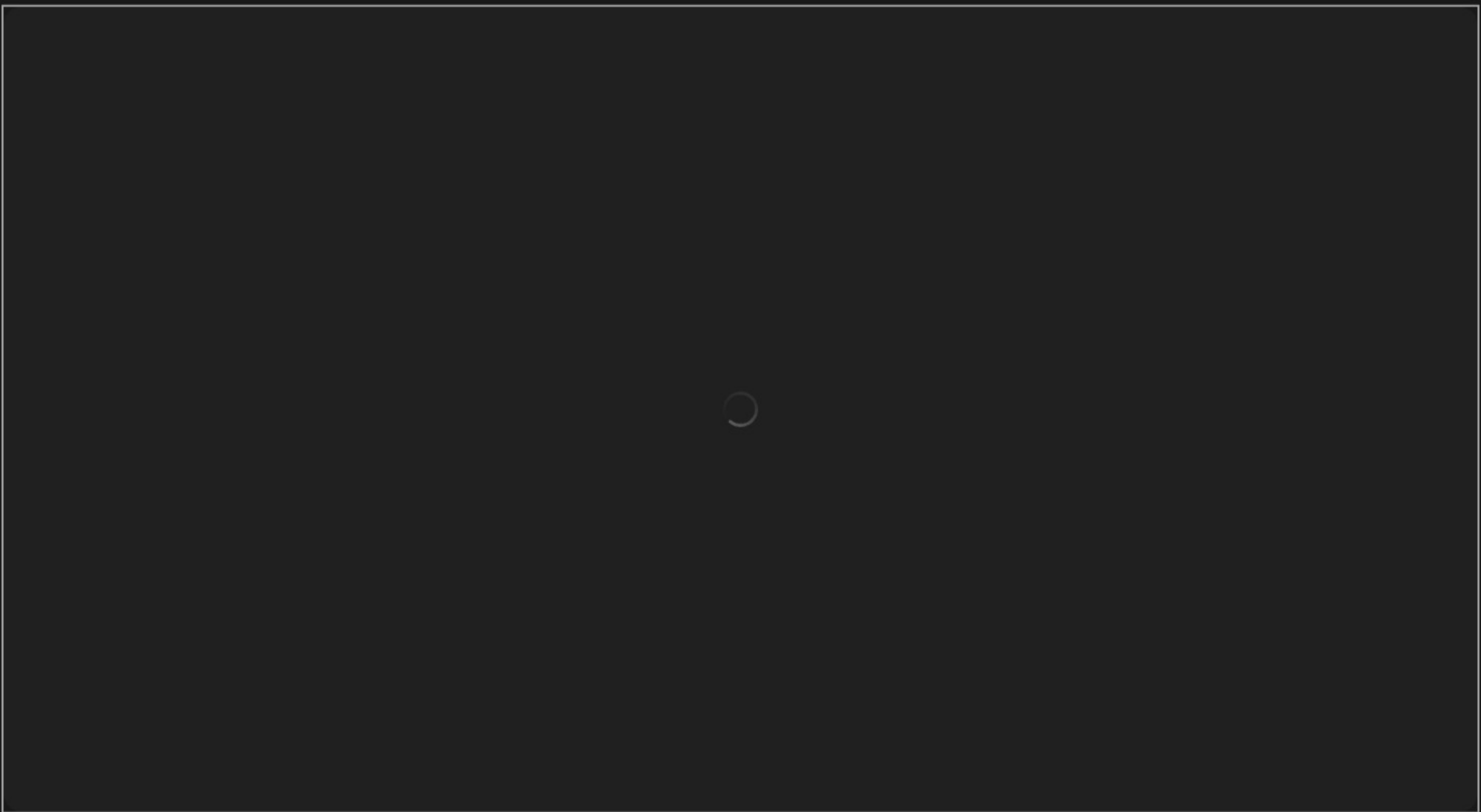
TiDB v6.5.6 起，Grafana 集成 DR–Auto–Sync 监控，展示集群同步状态。监控重点包括主副中心数据同步、Region Leader 和 PD Leader 分布，设置告警：集群状态异常、副中心 Leader 存在时触发告警，PD Leader 状态变化用于切换判断，确保运维及时响应。



◉ 注意事项 ◉

- **副本数目选择**：Raft 组成员数应为奇数以提高容灾能力，如三成员结构能容忍一成员故障，避免网络隔离导致服务不可用。
- **原生 Raft 局限性及扩展**：原生 Raft 在偶数可用区支持不友好，TiDB 通过扩展 Raft 功能，改善了在主可用区故障后，灾备区数据一致性恢复的难题。

◉ DR Auto–Sync 方案 ◉

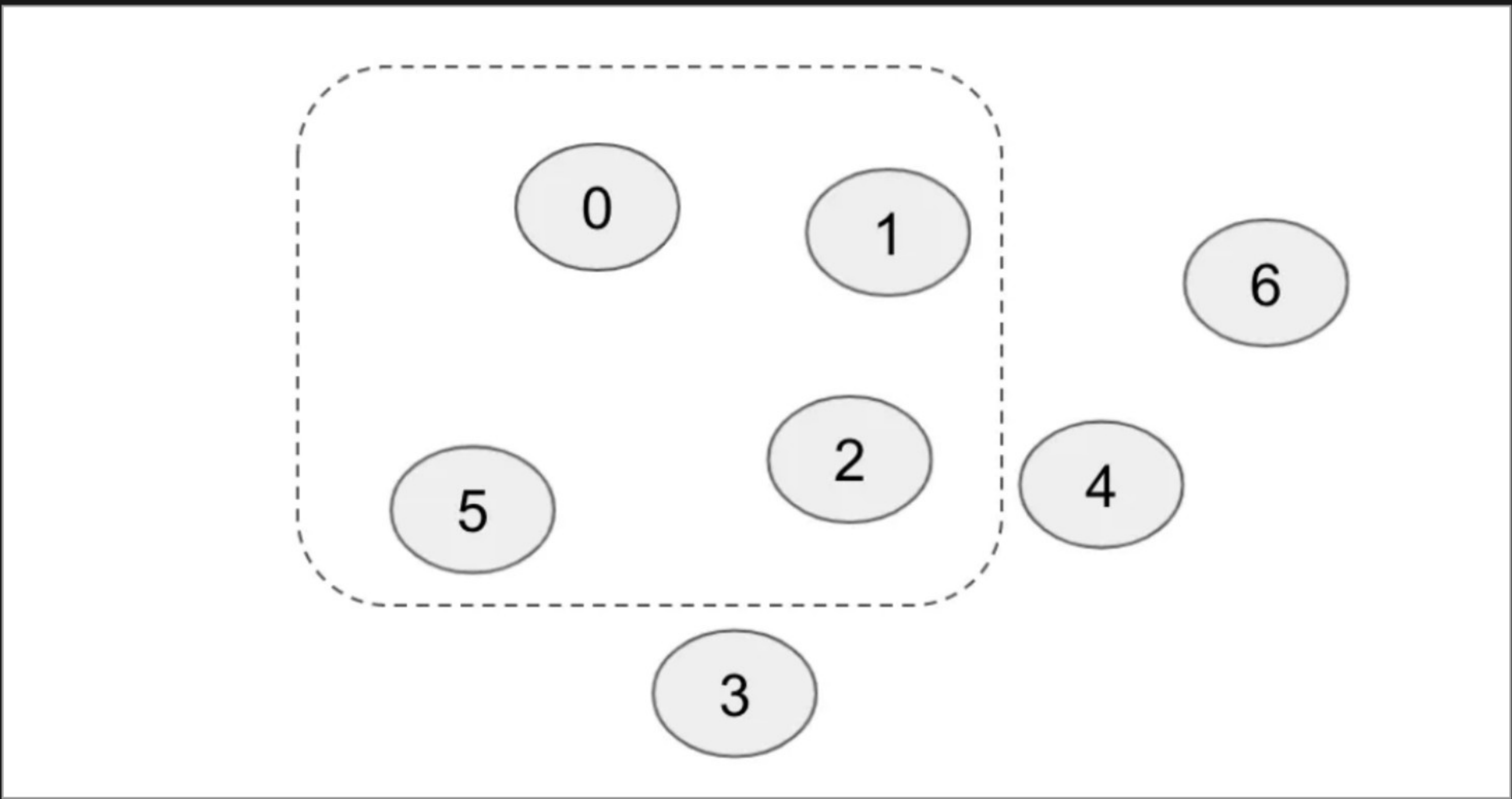


DR Auto-Sync 是 TiDB 的同城双中心数据同步方案，通过内部 Raft 协议实现两中心数据同步，保障任一中心故障时数据一致性。

- **同步状态**：TiDB DR-Auto-Sync 支持 sync、async、sync-recover 三种模式，自动调整保障数据一致性和业务可用。
- **状态转换**：PD 监控网络和中心健康，自动控制集群在 sync 和 async 间切换，确保数据同步和集群稳定。
- **RPO & RTO**：DR Auto-Sync 实现 RPO=0，最小化数据丢失。RTO 涵盖多种故障场景，影响集群恢复服务时间。


⋮ 原理解读 ⋮


- **核心技术① —— Raft Commit Group**：引入提交组概念，确保日志复制至少覆盖两个不同组，增强跨 AZ 的数据一致性与容灾能力。
- **核心技术② —— 请求阻塞窗口**：设置阻塞窗口以保持同步复制状态，防止网络故障导致的数据不一致。
- **核心技术③ —— 自适应复制状态切换**：设置阻塞窗口以保持同步复制状态，防止网络故障导致的数据不一致。
- **核心技术④ —— 少数派灾难恢复**：在主 AZ 故障时，通过 PD 手动指定灾备 AZ 的副本为 Leader，实现少数派副本的数据服务恢复。




⋮ 最佳实践 ⋮

- **同城双中心方案（DR Auto-Sync）**：

优点 ：实现同城双中心高可用和容灾，RPO = 0 数据恢复，自动切换异步复制，快速服务恢复

缺点 ：运维复杂，可能需人工介入数据恢复，跨 region 读写存在 ACID 问题。


● 同城三中心方案：

优点 ：三个数据中心提供高可用性和容灾，自动故障切换，无数据丢失

缺点 ：高网络稳定性要求，性能受网络延迟影响，写入和读取延迟增加

● 两地三中心方案：

优点 ：跨城高可用性，应对自然灾害，自动故障恢复，无数据丢失

缺点 ：受网络延迟影响性能，写入读取延迟，异地数据中心数据恢复可能不保证 RPO = 0。



[点击此处 | 阅读原文](#)



 立即咨询 TiDB 企业版 

