从单一到多活，麦当劳中国的数据库架构迁移实战

原创 凌敏 InfoQ 2025年04月16日 13:30 浙江

作者 | 凌敏

IT 基础设施改造的这把火，终于从互联网行业烧到了餐饮行业。

过去十余年，互联网行业通过 IT 基础设施的革新，实现了从单一数据库到多活数据库架构的跨越，显著提升了业务的高可用性和容灾能力。如今，餐饮行业也沿着这一路径，开始向多活数据库架构迁移。

从表面上看，餐饮作为最传统且极度下沉的行业，似乎和相对复杂的多活数据库架构关联不大。但实际上，随着餐饮企业数字化转型加速，从顾客点单到会员管理，从食材采购到营销推广，沉淀的数据资源早已达到海量规模。根据国家信息中心发布的《中国餐饮业数字化发展报告（2024）》，餐饮业通过智慧农业、数字采购、冷链运输等多环节的数据采集，形成了庞大的数据资源。报告强调，数据要素是推动餐饮业数字化发展的关键动力之一，具有“乘数效应”，能够有效促进数据要素自由流动，推动餐饮行业朝着智能化方向发展。

而多活数据库架构正是应对这一趋势的重要支撑。通过在不同数据中心部署多个数据库节点，能够确保即使任何一个数据中心故障，业务依旧可用。这种架构能够提升系统的高可用性和容灾能力，为餐饮企业的数字化转型提供了坚实的技术基础。

对于还没有采用多活数据库架构的餐饮企业而言，麦当劳中国的迁移实践提供了一个值得借鉴的范本——这个在国内坐拥 6000 余家门店的全球餐饮领先品牌，正通过 BCP（业务连续性计划）项目加速推动 IT 基础设施革新。其中，数据库作为确保业务连续性最关键的环节之一，从单一数据库架构迁移到多活数据库架构，自然成为了重中之重。

单一数据库架构，需要革新了

对于金融这类天然对高可用性有较高要求的行业而言，多活数据库架构早就成了必选项。但在拥有上千年历史的餐饮行业中，多活架构的应用实践并不普遍，餐饮企业更多会基于成本、技术复杂度、数据一致性等考量，选择采用单一数据库架构，即所有数据集中存储在一个数据库实例中。

当企业的业务体量和复杂度都相对较低时，单一数据库架构不失为一个性价比极高的选择。但当业务体量和复杂度陡增，单一数据库架构的局限性愈发明显。比如，当业务体量

增长到一定程度时，单一数据库的读写压力也会随之增大，容易出现性能瓶颈，进而影响系统响应速度。最致命的是，单一数据库架构等同于“把鸡蛋都放在一个篮子里”，一旦出现人为操作错误、电力系统故障、物理灾害等风险导致数据库宕机，整个业务系统也会跟着全面停摆。

“目前在餐饮行业中，多活架构的应用尚不普遍，企业更多是在数字化转型的推动下开始逐步考虑做多活。此外，多活架构在餐饮行业推广面临的挑战也比较多，包括技术要求较高、成本投入大以及数据一致性等问题。”尽管如此，麦当劳中国 IT 团队指出，**随着数字化转型的深入，多活架构正逐渐成为餐饮行业的一个确定趋势，特别是对于大型餐饮企业来说，构建多活的高可用架构已经成了保障业务持续稳定运行的必然选择。**

麦当劳中国正是基于业务稳定性与连续性的考量，对数据库进行了 3AZ 改造，即在三个可用区（Availability Zone，简称 AZ）部署数据库。

从理论上讲，像麦当劳中国这样的餐饮巨头，过去采用单一数据库架构并非最佳实践。但在麦当劳中国进行本地数字化转型的初期，为了简化系统复杂度，单一数据库架构自然成为一个过渡性的阶段性状态。此外，麦当劳中国当时还面临另外一个现实挑战：一个运行超过 10 年的数据中心也需要迁移到新的中台方案中。在完成整体迁移之前，系统需要暂时依赖单一数据库架构来维持正常运行。

2019 年，麦当劳中国开始正式推进数字化转型进程。随着企业 IT 服务不断增加，来自 APP、小程序的流量和订单占比越来越高。在用餐高峰时段，一旦 APP 或小程序出现故障导致服务中断，将会对业务产生较大影响。基于业务连续性的考量，2022 年，麦当劳中国启动了 BCP 项目，这也是麦当劳中国构建 IT 基础架构时非常重要的核心内容。而在 BCP 项目中，最核心的就是保障用户数据和业务数据的完整性——即便在机房级别故障的情况下，核心业务仍能正常运行。

麦当劳中国率先组建了一个由测试、研发和运维人员组成的 BCP 项目团队，而数据库正是整个项目最关键的一环。团队采用了 TiDB 分布式数据库，并投入了近一年的时间，对 TiDB 的 BCP 方案进行了全面而深入的调研，最终成功完成了 TiDB 的 3AZ 改造，将原有的单中心架构升级为三中心架构。

据麦当劳中国 IT 团队介绍，本次改造的范围包括麦当劳点餐主流程中的所有核心系统，如会员、积分、订单和卡券等。

通过监控数据发现，**改造后的数据库平均响应时间比单中心架构下减少了约 20%，系统性能提升了 30%-50%。此外，改造完成后，数据库集群的容灾能力大幅提升，业务恢复时间明显缩短。** 在过去的单中心架构下，如果发生机房级别的故障，业务将完全不可用；而在三中心架构下，即使任何一个机房发生故障，数据库可能在 10 分钟内就完成了业务恢复，整个业务系统在半小时内即可恢复正常服务，极大地提升了业务连续性。

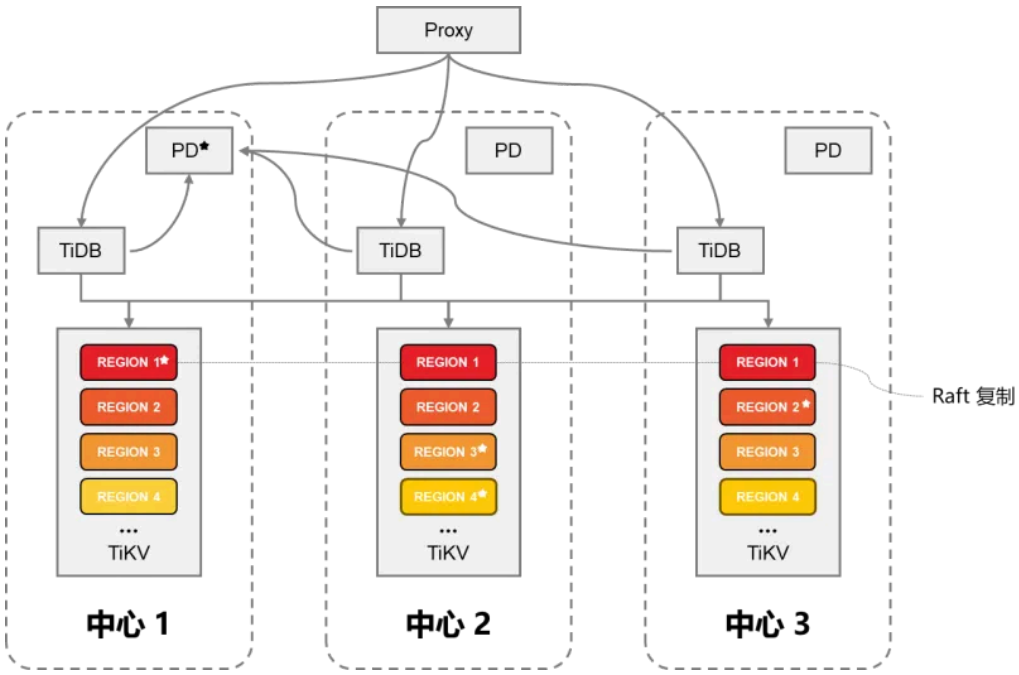
五花八门的架构方案，

应该怎么选？

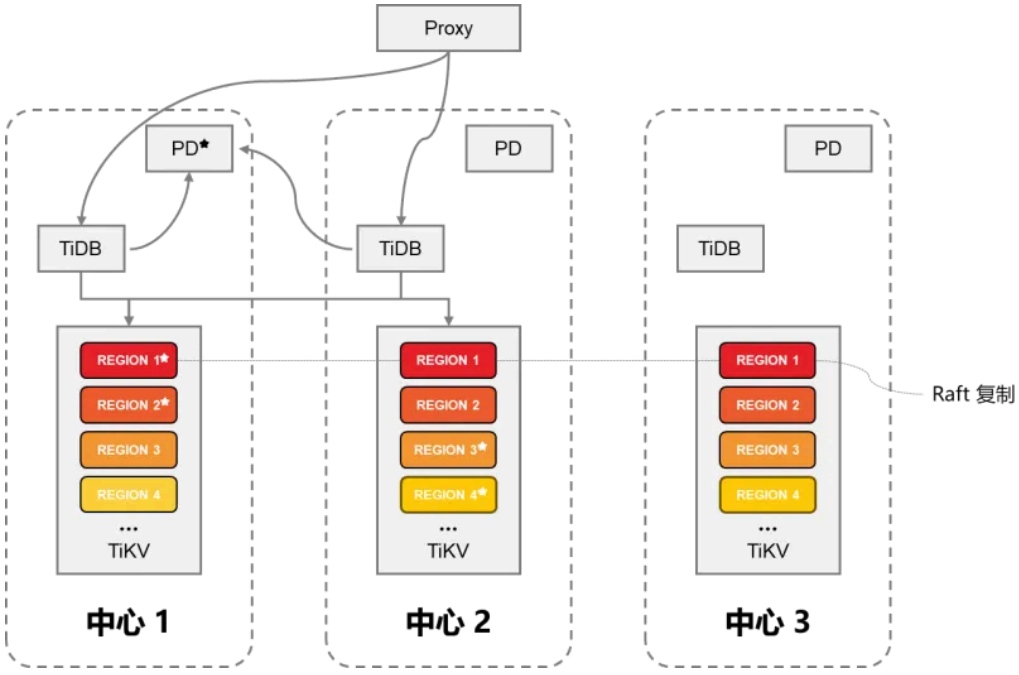
在多中心改造过程中，数据库的 CAP（一致性、可用性、分区容错性）是重中之重。 为确保找到更适合自身业务场景的技术方案，麦当劳中国的项目团队从 2022 年 5 月开始对 TiDB 的 BCP 方案进行调研，期间开展了多轮的测试和 POC(Proof of Concept，概念验证) 实践，并根据结果持续优化和调整方案。

在调研初期，团队重点考察了四种架构方案：

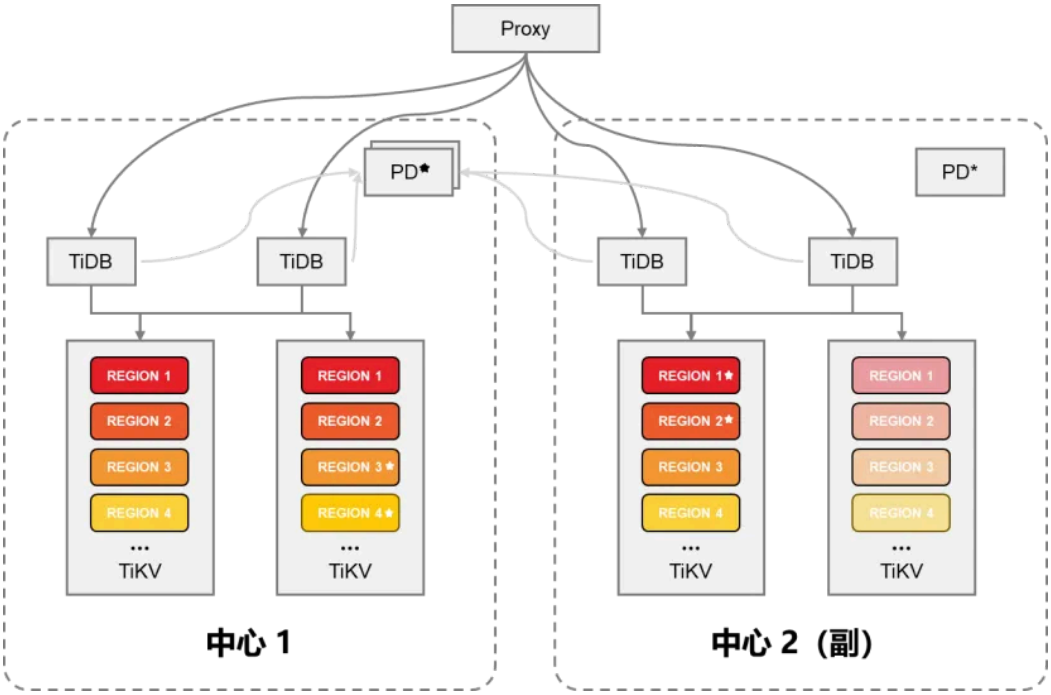
三中心单集群方案：TiDB 使用 Raft 协议作为共识机制，与三数据中心架构兼容性极佳。集群跨三个数据中心部署，每个中心都可对外提供读写服务。即使任一中心故障，系统仍能正常运行，且数据一致性不受影响。



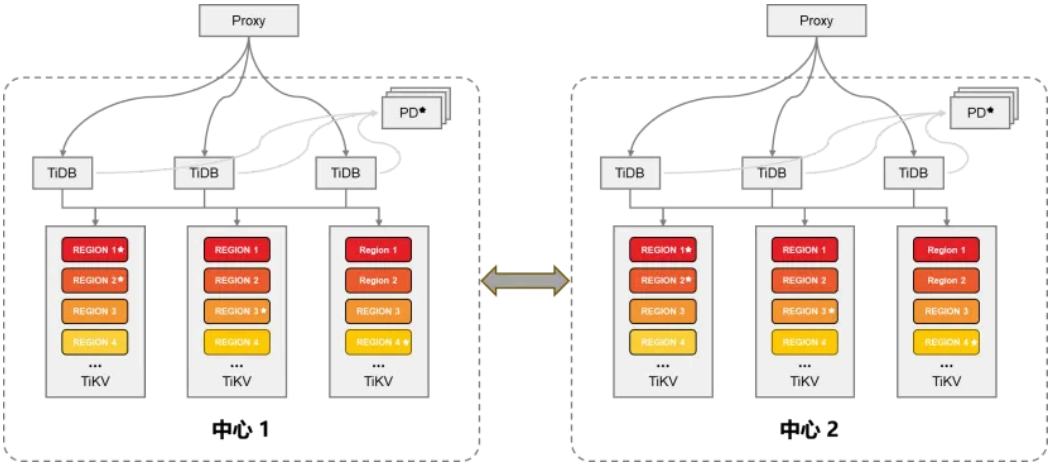
“伪”三中心单集群方案：这是三中心单集群方案的变种，第三个中心仅作为仲裁节点，不对外提供服务。实践中，前两个中心通常为 IDC 节点，第三个中心可以是低配置的云节点，从而降低网络成本和复杂性。



双中心单集群方案：TiDB 集群部署在两个数据中心，比三中心更经济。由于 Raft 依赖多数派共识，副本以 2:1 比例分布，主中心两份，副中心一份，外加一份无投票权副本。副本间通过 Raft 保证一致性和高可用。该方案有三种同步状态：同步复制（sync）、异步复制（async）和恢复同步（sync-recover）。



传统双中心互为主从方案：在双向复制容灾方案中，两个地理位置分散的 TiDB 集群互为数据的备份，仿照 MySQL 的双向主从复制。当一方故障时，另一方无缝接管，保障高可用性。



在调研过程中，团队发现双中心单集群方案由于 TiDB 生产版本（V4.x）的限制存在一些缺陷，因此首先排除了这一方案，重点对另外三个技术方案做对比和测试。

其中，双中心互为主从方案是 TiDB、MySQL 或其他关系型数据库做 BCP 和高可用架构设计时普遍采用的常规方案。在这种架构下，两个数据中心互为主从关系，一个作为主中心处理主要业务流量，另一个作为从中心进行数据备份和容灾准备。但这种方案的局限性在于，当网络延迟较高或复制链路出现问题时，可能存在数据延迟和一致性问题。且该方案需要对业务代码进行改造，保证数据不会在两个中心重复执行。因此，该方案 **更适用于那些对数据一致性要求较低的场景**，如社交媒体或电商等。

三中心单集群作为最主流的技术方案，通常采用三中心五副本架构，**适合对数据一致性要求极高、性能影响敏感的场景**，如银行交易系统等。但该方案对网络延迟和带宽的要求极高，成本也比较高。此外，考虑到麦当劳中国当时只有两个 IDC 数据中心，其他资源主要依赖于公有云服务，而 IDC 和公有云之间的专线质量相对不够稳定，这种限制促使团队选择了一种更灵活、更适应实际条件的架构方案——“伪”三中心单集群方案。

“伪三中心单集群是我们自己取的名字，并不是一个通用叫法。我们调研发现，这一方案更契合麦当劳中国的需求，可以通过两个 IDC 对外提供业务，另外一个云上的 IDC 只承接数据，不对外提供服务，从而降低了 IDC 与云之间的网络专线不稳定对性能的影响。”

据麦当劳中国 IT 团队介绍，在评估不同技术方案时，团队重点关注了两个关键指标：**RPO（恢复点目标）和 RTO（恢复时间目标）**。“我们要求 RPO 为 0，数据必须保证完全一致；RTO 为 10 分钟，发生故障后，系统必须在 10 分钟内完全恢复所有业务。”对比其他企业来看，麦当劳中国对 RPO 和 RTO 两个指标的要求都相对较高，其中 RPO 为 0 的要求更是与金融行业一致。

除了 RPO 和 RTO，团队在评估技术方案时还重点关注了网络质量、成本和运维复杂度等关键因素。

其中，网络延迟和带宽能够直接影响方案的可行性，因此网络质量成为团队首要考量的因素。调研发现，如果网络延迟超过 3 毫秒，双中心互为主从方案可能是更合适的选择；而如果延迟低于 3 毫秒，“伪”三中心单集群方案则更具优势。

成本也是团队关注的重点，具体包括硬件、软件以及数据丢失后可能带来的损失等。从成本角度来看，相比需要双倍服务器配置的互为主从方案，“伪”三中心单集群方案仅需 1.5 倍的服务器数量，即可将性能提升 30%-50%。此外，“伪”三中心单集群方案的运维复杂度相对较低，特别是在故障恢复时无需人为干预，配置也较为简单。

如何实现架构平滑迁移？

制定迁移策略

确定好“伪”三中心单集群方案后，下一步就是完善实施方案，这也是实现整个迁移过程更丝滑的基础。其中，最重要的就是制定迁移策略。

最初，项目团队考虑采用基于主从复制技术的主从灾备切换方案进行数据库架构变更。该方案将主集群的数据异步实时复制到从集群。迁移时，需停止主集群写入，等待数据同步完成并验证一致性，然后将应用连接切换至从集群。然而，停止写入、数据同步和一致性比对过程预计至少需要 10 分钟的停机时间，考虑到麦当劳 24 小时营业的特殊性，团队决定寻找一种不影响业务的实施方式。

经过与 TiDB 原厂、研发和测试团队的深入讨论，在缜密的分析和论证之后，团队最终选定了原**地扩缩容方案**（在线将集群从单中心部署变更为单集群三中心部署）作为本次多中心改造的核心策略。而对于数据量庞大的会员系统，团队采用**主从集群切流方案**。该方案通过搭建一套 3AZ 架构，将数据单向同步，可以将切换时间控制在 5 分钟以内。

在正式上线前，团队先针对每套集群在测试环境中开展了多轮原地升级演练和故障演练。在演练过程中，团队发现并解决了 30 个性能相关问题和 8 个故障演练相关问题，并逐一攻克了技术难点和疑点，确保了方案的可靠性。

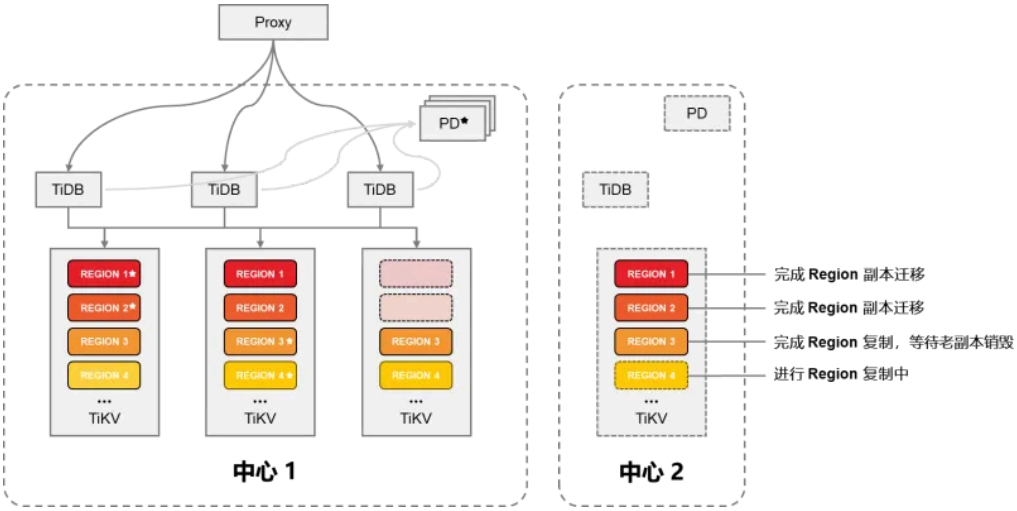
值得一提的是，第一套试点集群的测试周期长达 80 天，通过经验积累和持续优化，后续集群的准备时间缩短至约 1 个月。整个准备阶段的系统性和严谨性，为生产环境的平稳上线提供了有力保障。

实施改造

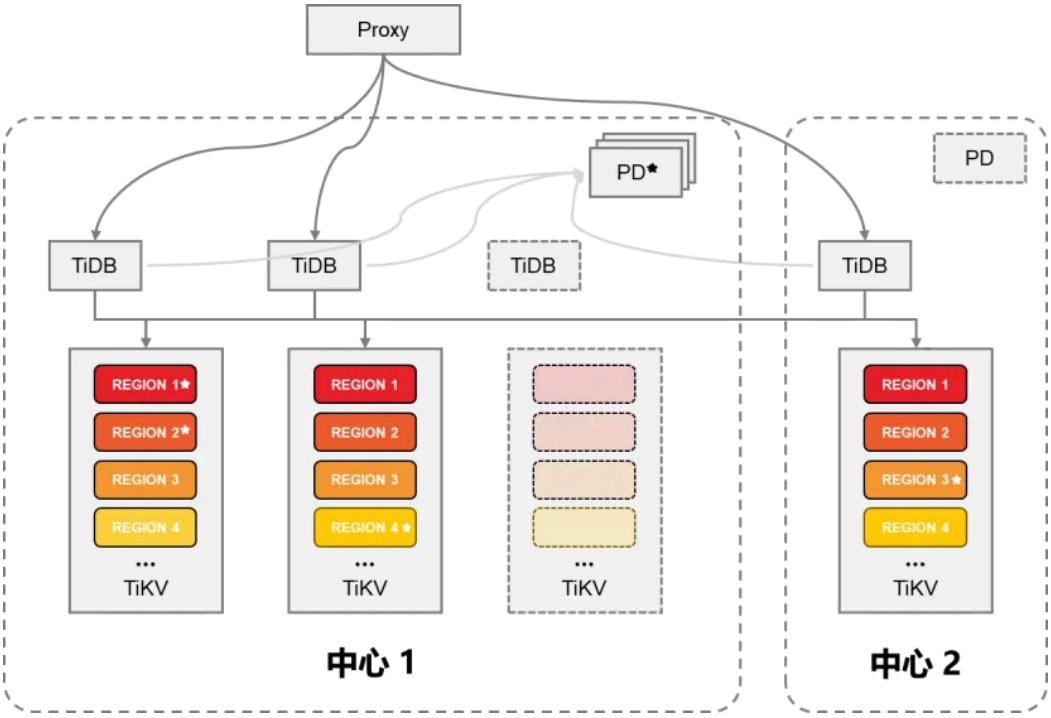
在生产环境的实施阶段，团队选择凌晨 0:00-04:00 这一业务低峰时段进行变更操作，最大限度地降低了迁移过程对业务的影响，实现用户无感知的平滑升级。为了确保上线过程万无一失，团队制定了详细的回滚方案，确保每个步骤都可以快速回滚，以便在出现问题时能够快速恢复，避免影响业务。

具体来说，团队将整个迁移过程划分为三个阶段逐步推进：

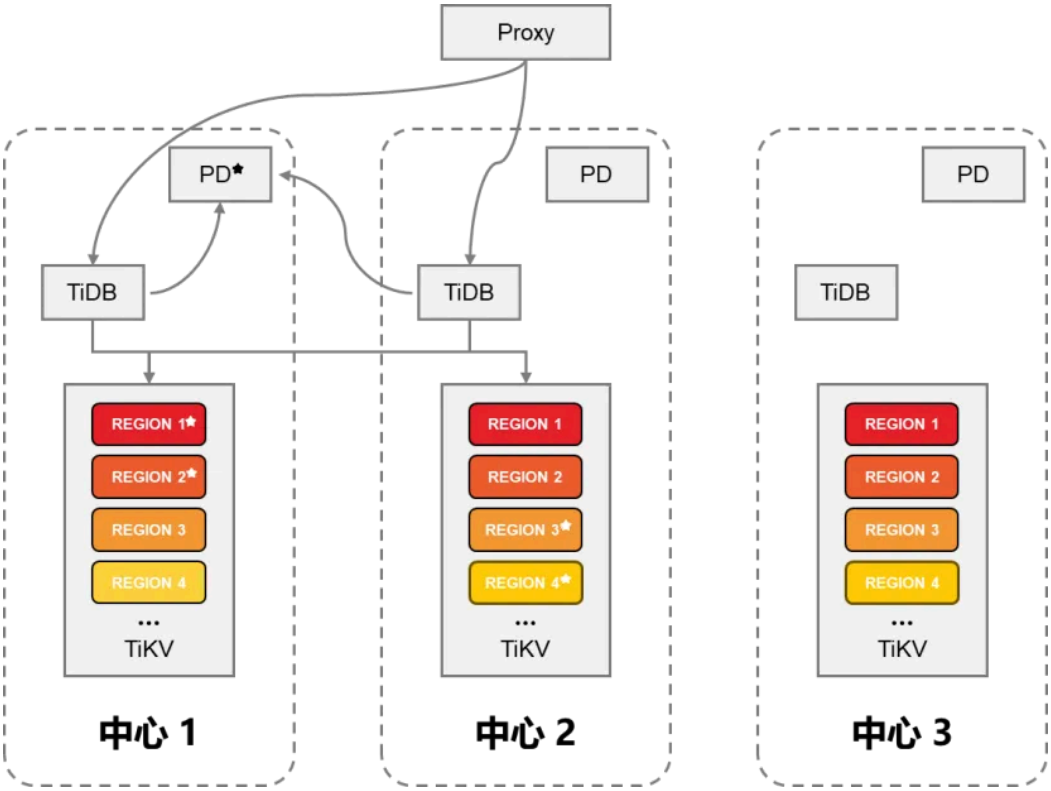
第一阶段： 在第 2 个中心部署 TiDB 实例，并将其加入原集群。借助 TiDB 的在线弹性扩容能力，系统自动以 Region 为单位逐一迁移数据副本。整个过程采用“小步快跑”的方式，人为控制迁移速度，并设置数据中心 2 的节点暂不承担业务流量，将迁移对业务的影响降到最低。



第二阶段： 当所有数据副本在数据中心 2 中生成后，原集群的部分节点退出，数据中心 2 开始承担业务负载。此时，如果遇到问题，团队可以快速回退至单中心运行，确保业务连续性。



第三阶段： 在确认数据中心 2 稳定运行后，团队重复类似第一阶段的操作，将数据中心 3 的节点以扩容的形式加入集群，TiDB 自动以 Region 为单位逐一迁移数据副本。整个过程同样采用“小步快跑”的方式，人为控制迁移速度，并设置数据中心 3 的节点暂不承担业务流量。



“实施改造阶段的技术复杂度最高，因为涉及到真实的生产环境，同时也是技术方案确定后需要进一步打磨的环境，所以当时遇到的问题也比较多。”在第二套集群的实施过程中，团队曾遇到过一个重大挑战：在扩缩容操作后的凌晨 4 点，集群的响应时间出现异常升高。团队迅速组织紧急会议，包括研发、测试、DBA 和 TiDB 原厂工程师共同排查问题。由于此前团队已经制定了详细的回滚方案，在接近业务早高峰时，团队决定如果问题还是无法

得到解决，将执行快速回滚。最终，团队在业务早高峰前成功定位到问题根源在于一个参数需要调整，并在生产环境中紧急修改。5 分钟后，集群的响应时间恢复到正常水平。

除了提前制定详细的回滚方案，团队还准备了一个兜底方案：搭建了一套与原单中心架构相同的集群。如果在三中心变更过程中出现问题，无法回滚到最初的单中心状态，可以通过这套备用集群进行紧急切换，确保业务不受影响。此外，在整个升级过程中，麦当劳中国项目团队和 TiDB 原厂的技术工程师始终保持在线，从开始变更操作到早高峰结束，全程监控业务运行情况，保障整个实施改造过程能够顺利进行。

对于这样一个大型跨团队协作项目来说，团队间的紧密配合与高效协同是确保项目顺利推进的核心因素之一。

据介绍，整个项目共涉及研发、测试、DBA、产品、业务、运维等多个团队，由性能测试团队的成员担任项目经理，通过内部自研 Ninja 工具进行任务调度和发布管理，确保项目能够高效执行。

在跨团队协作过程中，最大的挑战来自于沟通成本。不同团队的工作模式和技术能力存在差异，导致大家在规划时间和实施方案时，需要花费大量精力协调各方需求。为此，项目团队**建立了定期的沟通机制**，包括周会、月会以及关键问题讨论会，确保所有相关信息能够及时同步到各个业务方和跨团队部门。

在研发、测试和 DBA 等跨团队成员的紧密协作下，通过前期调研、测试演练和实施改造三个关键阶段的缜密执行，麦当劳中国圆满完成了 TiDB 多中心改造，不仅提升了系统的可靠性和容错能力，为业务连续性保障奠定了坚实的基础，也为未来的系统设计和改造提供了可复用的方法论。

写在最后

在完成多中心架构迁移后，麦当劳中国对多活架构的未来优化方向也有了更清晰的规划。“我们现在实现的是同城双活，未来的目标是朝着多地多活架构演进，不仅局限于现有的下单链路，还将扩展到更多的业务场景，例如 ToB 链路的多活支持。”

对于其他考虑从单一数据库架构迁移到多活数据库架构的企业，麦当劳中国也分享了以下几点建议和经验：

1. **从业务数据出发，明确需求**：企业在设计多活架构时，首先需要根据自身的业务特点和数据重要性明确需求。例如，麦当劳中国要求数据零丢失（RPO=0），这对架构设计和成本提出了更高要求。如果某些业务可以接受一定程度的数据丢失，方案选择可能会更加灵活，成本也会相对降低。
2. **提前规划多活数据中心**：多活架构的部署需要提前规划，尤其是对数据中心的布局和网络质量的评估，网络延迟和带宽直接影响架构的可行性和性能表现，因此企业需要在早期阶段就对网络条件进行充分测试和优化。
3. **建立高效的跨部门合作机制**：多活架构的迁移涉及到多个部门的协作，企业需要建立高效的跨部门合作机制，确保项目能够顺利推进。