

聊聊跨数据库迁移的数据比对那些事儿

原创

吾亦可往

2025-03-06



309

各位技术小伙伴们，今天来和大家深入探讨一下跨数据库迁移过程中至关重要的数据比对环节。在当下这个信息数据呈爆炸式增长、数据库技术不断革新的大环境下，数据库迁移已经变得越来越普遍。而数据比对作为保障迁移准确性与完整性的关键步骤，其重要性怎么强调都不为过。

一、现有数据比对技术的困境

目前，国内外关系型数据库数据比对技术一直在持续发展，但现存的方法普遍存在不少问题。以全量比对为例，许多常规做法是导出两端表的全列数据，或者仅导出源端表的全列数据来实施比对操作。然而，这种方式会导致导出的数据文件体量极为庞大，解析过程异常缓慢，效率极其低下。更为关键的是，这种方法完全没有充分考量不同类型表在比对时所需的时间成本差异。

再看增量比对，由于增量数据本身具有动态变化的特性，在导出增量数据的过程中，根本无法确切确定这些数据的实时有效性。这就直接导致最终的比对结果准确性大打折扣，难以满足实际业务对于数据精准度的严格要求。那么，面对这些困境，是否存在更为有效的解决办法呢？答案是肯定的，接下来要为大家介绍一种新的跨数据库迁移数据比对方法，十分值得我们深入研究与探讨。

二、新的数据比对方法详解

（一）数据拉取

这种新方法的第一步，是借助大数据组件快速、高效地拉取待校验数据。大数据组件在此过程中发挥着关键作用，它能够从复杂的数据环境中精准地获取所需数据，为后续的数据处理流程奠定基础。

（二）数据分类

拉取到数据后，紧接着便是一个极为关键的步骤 —— 将待校验数据按照表类型进行细致分类。具体的分类方式如下：首先对数据进行解析，获取数据关键列的内容。若关键列内容不为空，那么该数据类型大概率属于主键表或唯一索引表；若关键列内容为空，则需进一步判断数据是否含有普通索引。若不存在普通索引，此数据类型即为三无表；若含有普通索引，则可判定为普通索引表。通过这种严谨的分类方式，能够为后续的查询操作提供更为精准的依据。

（三）查询操作

完成数据分类后，接下来要根据不同的表类型获取与之对应的查询条件。对于主键表和唯一索引表，以主键列作为查询条件；三无表由于其特殊性质，以全字段作为查询条件；普通索引表则以源端索引列为查询条件。有了明确的查询条件后，就可以拼接对应的查询语句，并在数据库中执行查询操作。通过这种针对性的查询策略，能够显著缩短查询时间，提高数据处理效率。

（四）数据比对与记录



吾亦可往

关注

455

文章

169

粉丝

148K+

浏览量

- 获得了 1759 次点赞
- 内容获得 960 条评论
- 获得了 644 次收藏

TA的专栏

- GoldenDB

goldendb专利

收录 112 篇内容
- GoldenDB

GoldenDB文档

收录 4 篇内容
- GoldenDB

GoldenDB论文

收录 12 篇内容
- < >

热门文章

- GoldenDB安装

2024-03-19

4224浏览
- GoldenDB去O新进程，DSG助力某华东省移动账管系统自主可控

2023-08-07

4208浏览
- GoldenDB导入导出使用介绍

2024-04-11

3663浏览
- 焕新出发，以新求变——广发银行成功上线分布式银行核心系统

2023-09-04

3139浏览
- 手把手安装搭建GoldenDB

2023-05-22

3049浏览

最新文章

- GoldenDB：分布式数据库系统中随机数一致性保障的创新探索

10分钟前

1浏览
- GoldenDB 数据查询优化：创新方法提升性能

23小时前

15浏览
- GoldenDB 在分布式数据库实时数据处理中的卓越实践与创新探索

4天前

36浏览
- 深度剖析 GoldenDB 在分布式数据库变动数据捕获中的创新实践

4天前

31浏览
- GoldenDB：引领数据库业务测试变革，提升金融科技效能

6天前

30浏览

查询操作完成并得到结果后，便进入以表为单位的数据逐级比对环节。在这一过程中，系统会逐列对数据进行细致比对，优先校验源端数据库和目标端数据中对应主键列的存在情况。若主键列不存在，则采用其他数据进行列比对。经过比对，差异化数据便会浮出水面。此时，将这些差异化数据上报至大数据组件，并进行落盘记录。这一举措具有重要意义，它使得后续查看比对结果变得更加便捷，极大地增加了数据比对结果的可视性，方便技术人员对数据差异进行深入分析与处理。

（五）数据修复

最后，根据全量数据和增量数据的不同性质，通过预设的修正策略对差异化数据进行修复。在全量阶段，一旦存在差异化数据且数据比对完成，系统便会发送修复指令；在增量阶段，当上游处于停机状态时，同样会发送修复指令。一旦系统响应修复指令，会立即停止拉取待校验数据和数据比对操作，集中精力对差异化数据进行修复。修复过程以表级为单位，以源端数据为基准，执行 delete 和 insert 操作，确保数据库迁移前后的数据一致性。

三、新方法的显著优势

（一）灵活性大幅提升

这种新方法针对不同类型的表分别制定了独特的比对处理策略，完全摆脱了对中间数据文件的依赖，极大地提高了数据比对的灵活性。无论是面对复杂的表结构，还是多样化的数据类型，都能够游刃有余地进行处理。

（二）性能优化显著

该方法创新性地提供了列级数据比对方式，并支持多线程并发执行。这一技术手段使得数据比对的性能得到了大幅提升，能够在更短的时间内处理海量数据，满足企业对于高效数据处理的迫切需求。

（三）强大的映射支持

它支持库表、一对一、一对多列级映射，同时还具备忽略列比对的功能。这种全面而强大的映射支持能力，使得在处理各种复杂的数据迁移场景时，都能够确保数据比对的准确性与完整性，为企业的数据迁移工作提供了坚实的技术保障。

（四）多样化的校正方法

新方法还提供了不同的校正方法来修复数据。这些多样化的校正方法能够根据不同的数据问题和业务需求，灵活选择最合适的修复方案，确保数据的准确性和一致性，有效降低数据迁移过程中的风险。

四、总结与展望

在跨数据库迁移的复杂工作中，选择正确的数据比对方法无疑是重中之重。上述介绍的新方法在数据比对的灵活性、可视性以及性能等多个关键方面都实现了重大突破，为我们解决跨数据库迁移的数据比对难题提供了全新的思路与有效的解决方案。

希望大家在实际工作中，如果有用到类似的技术，或者遇到相关问题，都能在这个论坛里积极分享与交流。通过大家的思想碰撞，我们或许能够激发出更多创新的技术火花，共同推动数据库技术在跨数据库迁移领域的不断进步与发展。期待大家的精彩分享！

 大数据 数据迁移 数据库 goldendb

「喜欢这篇文章，您的关注和赞赏是给作者最好的鼓励」

关注作者

赞赏

【版权声明】本文为墨天轮用户原创内容，转载时必须标注文章的来源（墨天轮），文章链接，文章作者等基本信息，否则作者和墨天轮有权追究责任。如果您发现墨天轮中有涉嫌抄袭或者侵权的内容，欢迎发送邮件至：contact@modb.pro进行举报，并提供相关证据，一经查实，墨天轮将立刻删除相关内容。



评论

分享你的看法，一起交流吧~

 luyingjun 

聊聊跨数据库迁移的数据比对那些事儿

4月前 1 评论

 鲁鲁 

聊聊跨数据库迁移的数据比对那些事儿

4月前 1 评论

相关阅读

【DBA坦白局】第三期：作为DBA，你加过最晚的班是到几点？在干什么？

墨天轮编辑部 1401次阅读 2025-07-15 10:28:44

2025年7月国产数据库大事记：GoldenDB创千万级大单，可信数据库大会召开，openGauss HyBench打榜第一，电科金仓举办2025产品发布会.....

墨天轮编辑部 790次阅读 2025-08-05 17:20:33

2025年8月中国数据库排行榜：双星竞入三甲榜，TDSQL 连跃位次升

墨天轮编辑部 718次阅读 2025-08-07 16:22:24

优炫数据库在山东省寿光市人民检察院成功应用！

优炫软件 647次阅读 2025-07-16 09:50:01

重磅 | 万里数据库GreatDB亮相上合组织数字经济论坛 以硬核科技共绘“数字丝路”新图景

万里数据库 592次阅读 2025-07-15 09:45:41

2025年7月国产数据库中中标情况一览：长沙银行千万采购GoldenDB，秦皇岛银行近七百万采购TDSQL！

通讯员 587次阅读 2025-08-07 10:18:23

中国信通院2025上半年“可信数据库”新增标准解读

大数据技术标准推进委员会 562次阅读 2025-07-21 10:45:15

希望和国产数据库厂商共建公共知识库

白鳍的洞穴 480次阅读 2025-07-28 12:24:07

ACDU周度精选 | 本周数据库圈热点 + 技术干货分享（2025/7/25期）

墨天轮小助手 470次阅读 2025-07-25 15:54:18

ACDU周度精选 | 本周数据库圈热点 + 技术干货分享（2025/7/17期）

墨天轮小助手 436次阅读 2025-07-17 15:31:18