

电网数据中心数据同步策略的研究与设计

赵小凡, 杜舒明, 梁雪青

(广东电网有限责任公司广州供电局, 广东 广州 510000)

摘要:本文对电网数据中心数据同步传输需求进行了多场景的分析,并从全局的角度,在管理上设计了数据中心数据同步传输管理流程;在技术上针对不同的数据同步的场景,将数据传输场景进行2种归类,并设计了与之匹配的数据传输策略。在满足需求前提下,保证数据中心以高性能、低消耗、高质量的数据同步传输服务,为电网信息数据中心建设提供一个成功的案例和具体的解决方案。

关键词:数据中心;数据同步;数据服务;生产者/消费者模式

中图分类号:G623.58

文献标识码:A

文章编号:1004-7344(2021)31-0086-02

0 引言

随着数字化转型和数字电网建设工作的全面启动,电网数据中心从“旁站式”向“底座式”模式升级,成为数字电网建设的基础数据底座。

数据中心作为释放大数据价值的基础,是构建数据价值链的“连接器”,承载着数据汇聚、存储、共享分发的责任,而数据的汇聚、存储、共享分发,都离不开数据的传输与同步。传输与同步效率时刻影响着数据中心的数据服务质量。然而在当前的数据中心建设过程中,对数据同步问题重视不足、技术研究深度不够等因素,导致当前数据同步仅仅满足数据 T+N 的数据同步需求,而时效性要求应用场景暂时无法满足。

本文研究数据中心数据传输同步策略,为数据中心建设高效率、高时效的数据传输通道提供参考方案,助力数据中心释放大数据价值。

1 同步需求分析

首先,我们需要对数据中心的数据种类、数据同步需求和数据传输过程进行综合分析,以设计符合实际数据现状和需求情况的数据同步传输策略。

1.1 数据分类

数据中心是全网数据的汇集中心,数据类型多样,包含:TP/AP型数据(关系数据库的结构化数据)、时序型数据(日志、物联网数据)、半结构化数据(CSV、XML、JSON)、非结构化数据(电子邮件、文档、pdf)、二进制数据(图像、音频、视频)。

1.2 场景分析

数据中心的数据传输需求主要有以下四种场景:

1.2.1 事务数据交互

事务数据交互发生在多个业务系统之间需要数据交互的情况。一般情况下,事务型数据交互的数据交互数据量小、准确性、

时效性要求高。如营销客服接到电力扩建工程迟迟不完工影响出行的投诉,为了安抚用户,客服人员需要查询电力业扩业务系统,获取相应的工程数据,就需要向数据中心申请相关数据。

1.2.2 分析数据交互

分析数据同步需求,一般发生在跨业务的统计报表分析数据应用内。如企业发展部的报表数据应用通常会将生产、经营、营销、财务的重要关键指标汇总成为数据报表,以分析企业经营情况做决策与规划。此类交互的数据量大小由企业内部信息化程度决定,若各业务线的指标均已在各自业务系统内固化,则数据同步量会较小。若相关指标需要从业务数据明细里再重新汇总计算,则需要将需求范围内的业务数据进行全量同步,交互的数据量大。在时效性方面,实时运营决策的分析应用对数据时效性要求很高,汇报复盘类分析应用时效性要求较低。

1.2.3 全量同步

全量型同步需求,通常发生在数据中心汇总数据的过程。如数据中心将营销系统所有数据同步到数据中心的过程。

1.2.4 实时变更同步

实时变更数据同步需求,是指对数据变更敏感度高的数据同步需求,通常数据量较小,实时性要求较高。如重点数据运营监控等应用中,对变压器电力设备的负载信息监控信息会很敏感,以防负载过高导致供电故障事故的发生。

1.3 传输过程

数据中心的承载数据汇聚和共享分发的职能,对应的数据流向也有汇聚和分发,其数据传输流程如图1所示。

2 数据传输策略

2.1 设计原则

为达到高性能、低消耗、高服务质量的数据同步传输,我们遵循以下设计原则:

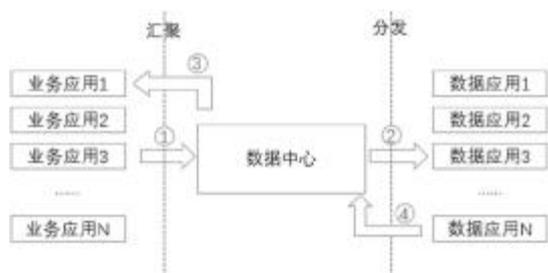


图1 数据传输

2.1.1 全局统管

从数据中心全局的角度出发统一管理,建立端到端的管理流程,保证每一个数据传输的需求与服务可看、可控、可管。

2.1.2 满足需求前提

以满足业务需求为前提下,从数据传输业务要求、数据类型、数据大小、数据时效性等方面入手全面分析业务需求,设计最优传输方案。

2.1.3 减少数据搬运

提前在数据中心进行指标汇总或计算,避免为了某些少量指标的计算,进行大量的数据搬运,占用存储、带宽资源的同时,数据搬运会带来全量数据泄露的安全风险。

2.1.4 沉淀服务积累

逐步沉淀数据中心提供的数据服务于 API,建立共享服务目录。减少冗余与重复建设,优化数据中心资源利用率。

2.2 全量同步

全量同步的典型场景是在数据中心将数据分发给应用的过程中,与源系统到数据中心 1:1 的同步过程不一样的是,数据中心支撑多个应用是 1:N 的关系,存在以下问题:

- (1) 1:N 的关系会产 N 条的同步通道,资源消耗成本大、维护困难。
- (2) 方案不稳定,通道随着应用的增多而增加,最终资源消耗殆尽。
- (3) 某些场景下应用使用的数据范围只是在某个范围内的子集,以全量的传输会造成资源的浪费。比如 2020 年新冠疫情电力复工指数分析应用,其数据范围仅仅是用电数据在 2018—2020 年间。

因此我们通过图 2 策略来避免以上情况:

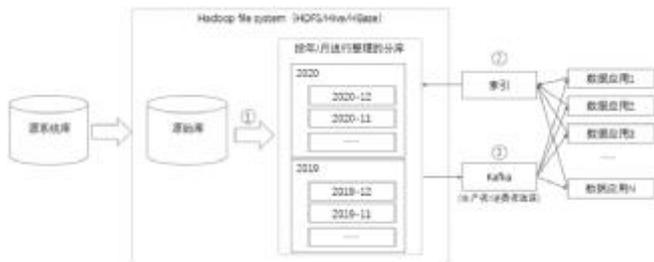


图2 数据同步

策略通过 3 个关键方式来满足 1:N 的数据同步需求:

- (1) 将贴源库进行年月维度的分拆,以分表的形式存储于另一个备份。
- (2) 建立分表目录与索引,数据应用根据自身的数据集合需

求,按年/月维度进行数据索引。

(3) 建立生产者/消费者消息通道,当数据发生变更时,生产者发送变更通知,消费者接受通知根据数据索引更新相关数据,保持数据的实时同步。

策略只需要维护好数据目录与消息通道即可,并不会因为应用的增多而带来更多的成本。并且对数据进行了分库处理,同步的量大小由应用自定义,减少了数据搬运。

2.3 分析型数据同步

同步发生在跨业务的统计报表分析数据应用内,存在数据需求量少、时效性高、通常以 API、共享服务等方式提供的共性,所以将其分为同一分类,使用统一策略进行数据传输服务,如图 3 所示,采用以下策略进行分析型数据的同步:

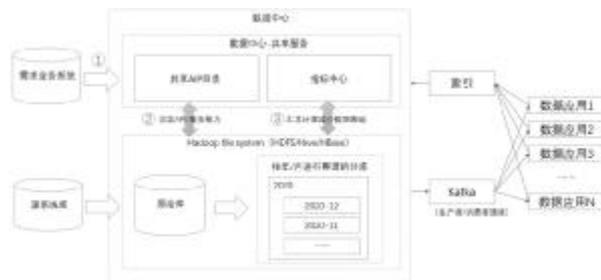


图3 数据共享

跨业务数据查询的时候,需求方向数据中心的共享服务提交数据请求,若数据中心已有相关数据的 API,则直接调用。若中心对应接口 API 未开发,则进行开发,完成后收录到 API 目录内,形成能力沉淀。

利用数据中心运算能力,提前将指标计算、数据汇总等工作完成,提供给数据应用直接使用,避免中心与数据应用之间的数据搬运,且可保证指标统一出口,保证数据的一致性。

2.4 策略优势

管理上提供需求分类和预判,根据不同需求选择不同数据同步策略,避免数据的乱传、乱接现象,保证数据传输需求与服务可看、可控、可管。

技术上,通过数据分表存储、目录索引建设,提前汇总数据、提前指标计算等工作,适当减少数据同步传输量,降低中心带宽,提升数据传输效率。

通过生产者/消费者模式,在数据中心和应用间建立消息通知渠道,利用数据变更消息由生产者(数据变动方)主动通知消费者(数据消费方)即时更新,从而保证数据的实时同步。

参考文献

- [1] 童述学.大数据中心基础设施建设[J].商品与质量,2019(26):59.
- [2] 杨娟娟.大型数据中心数据清理的策略与方法[J].中国金融电脑,2011(9):43-47.

收稿日期:2021-07-17

作者简介:赵小凡(1986—),女,汉族,湖北枣阳人,硕士研究生,高级工程师,主要从事信息技术、项目管理相关工作。