

企业信息系统数据交换平台设计与实现研究

乔 敏

零八一电子集团有限公司, 四川 成都 611700

[摘要]企业在信息化过程中存在的不同系统的数据无法互相访问的问题,本文设计了一个企业的信息管理系统之间的数据交换中心。通过对企业的数据交换的需求进行研究,着重对数据的采集、转换、传输、同步以及质量管理等方面的核心技术进行了研究,在此基础上完成整个平台的设计以及实现工作并对整个系统进行了测试与性能分析。经过验证可以有效的解决各个不同系统的之间的数据互通问题,使企业之间信息交换快捷、可靠、准确的完成。

[关键词]数据交换平台; 异构系统; 消息队列; 数据同步

DOI: 10.33142/sca.v9i3.19365

中图分类号: TN9

文献标识码: A

Research on the Design and Implementation of Enterprise Information System Data Exchange Platform

QIAO Min

Lingbayi Electronics Group Co., Ltd., Chengdu, Sichuan, 611700, China

Abstract: In the process of informatization, enterprises face the problem of data from different systems being unable to access each other. This article designs a data exchange center between enterprise information management systems. By studying the data exchange needs of enterprises, we focused on the core technologies of data collection, conversion, transmission, synchronization, and quality management. Based on this, we completed the design and implementation of the entire platform and conducted testing and performance analysis of the entire system. After verification, it can effectively solve the problem of data exchange between different systems, enabling fast, reliable, and accurate information exchange between enterprises.

Keywords: data exchange platform; heterogeneous systems; message queue; data synchronization

引言

伴随着企业的信息化发展,公司内部往往会部署着几个相互隔离的不同信息系统,比如ERP系统、CRM系统以及SCM系统等,在企业发展过程中各个阶段引入不同的厂商开发而成,技术体系,数据模型及存储方式各不相同。因为缺少统一的标准导致各个系统的间无法进行有效的交流,形成各自孤立的信息系统从而大大降低了公司的业务流转速度及决策水平。建立一套统一、快速稳定可靠的数据交换体系,达到不同系统之间的数据共享和业务协同是企业数字化转型过程中所面临的迫切需求。本文的研究目标是开发及实现一个企业信息系统的数据库、web服务接口、文件系统等不同类型的数据库、web服务接口、文件系统等不同类型的数据库,并要有可配置的数据转化映射能力来应对不同来源之间的数据模型及语义上的差别。对数据传送的要求是具有较高可靠性和较高吞吐率,保证传送中的数据不会遗漏也不会重复。在非功能性需求上,对实时性的要求也不尽相同,一些应用比如订单同步就需要做

明确企业的本质对于数据交互的需求是什么。从功能的角度来讲各个部门之间相互影响很大,一个部门做出的决策会影响到其他部门的相关工作,例如:销售部门下达了订单就需要同步到生产和库存管理系统中去,采购入库记录要实时传递给财务系统等,在不同系统的来回传送就是该平台建立的基础。目前企业存在的主要难题就是各个系统的数据标准不一样,接口协议也不一样,传统的一对一接口的的方式随着系统越多接口就会呈几何倍的增长,就像蜘蛛网那样复杂,管理和维护成本非常高昂而且不容易控制。从功能性的需求来说,平台应该有良好的多源异类数据集成能力,可以集成关系型数据库、web服务接口、文件系统等不同类型的数据库,并要有可配置的数据转化映射能力来应对不同来源之间的数据模型及语义上的差别。对数据传送的要求是具有较高可靠性和较高吞吐率,保证传送中的数据不会遗漏也不会重复。在非功能性需求上,对实时性的要求也不尽相同,一些应用比如订单同步就需要做

1 企业信息系统数据交换需求分析

企业信息化系统之间的数据交换平台的设计首先要

到准实时的数据更新，但报表数据可以进行批量的传送。

2 企业信息系统数据交换平台关键技术设计

2.1 数据采集与接入技术

数据采集及接入是数据交换平台的基本工作部分，主要任务是对各种异构的数据源进行快速集成并且抽取相应数据。对于企业内部的种类不同的数据源，数据交换平台采取插件式的适配器结构设计，每类不同的数据源对应一个专门的接入插件。针对关系型数据库数据源，使用基于日志的方式进行增量式的数据采集并通过对数据库事务日志来读取变化的内容，这种方法几乎不会给源系统造成负担，可以做到毫秒级的即时数据抓取；而对于那些不支持日志解析的数据库则使用基于时间戳或者事件驱动的方式来进行增量抽取^[1]。针对 web 服务接口，平台提供了统一的 HTTP/HTTPS 访问接口，同时支持 RESTful 以及 SOAP 这两种主要的标准风格类型，在数据抓取方面变更数据捕捉技术用于实时性强的数据聚合场合，在监控数据库源发生改变的操作仅获取变化后的新数据的方式可以大大的减少传输的数据量以及系统的压力，是实现实时数据同步的一种重要方法。

2.2 数据转换与映射技术

异质系统的不同数据形式是数据交互所遇到的主要难点，数据转换及映射方法就是应对这个问题的办法。平台采取可视化的配置式映射方式，在 GUI 上设置好从源到目标字段之间的映射关系。转换器可以进行很多种类型之间进行转化，例如常见的：字符串到数字、或者数字到字符串等简单的类型转换，也可以做复杂的类型转变，例如 JSON 对象到 XML 节点的转换。映射规则可以使用表达式来计算，可以自定义一些规则去拼接或者截取或者判断源数据等各种复杂的操作。数据映射难度直接影响着数据集成项目的执行速度及后续维护工作的难易程度，基于元数据驱动的数据映射方式通过对映射关系的抽离使转换规则脱离了代码逻辑，从而做到转换规则和代码相分离，可以由企业相关人员根据需要在界面上进行设置更新转换规则而不必改动程序代码，对于表与表之间的“一对多”或者“多对一”关联情况还支持嵌套映射功能以及父子级关联映射的能力，可用于解决主从表间的数据一致性问题。

2.3 数据传输与消息队列技术

数据传送安全性和传输带宽直接影响到数据交换系统服务质量。此平台以消息中间件为核心传输组件，借由消息中间件异步松散以及削峰填谷的优势解决企业间数据交换过程中不稳定的数据流问题。消息中间件的选择方面，根据吞吐率、延迟时间、消息持久化等条件挑选一款

适合做核心业务链路的消息中间件产品，以及一款适合处理高吞吐率日志类数据的消息中间件产品。消息队列为分布式系统重要基础组件之一，把消息生产者和消息消费者的分离，很好的把上游或者下游系统的依赖关系分离开来，在下游系统发生异常状态下消息队列可以暂存数据等待恢复之后再重新推送，彻底解决了消息传递时的数据可靠性问题，在消息传递可靠性的设计上，平台使用的是生产者提交机制，即保证消息写入到消息队列内才会进行返回成功的操作，在消费端使用的是手工应答的方式只有在业务执行成功的情况下才会给消息队列进行响应，如果失败的话会再次进入到消费队列中。对于防止因相同消息被消费引发的信息重复现象，在平台上增加在客户端的幂等处理策略来保存已经被消费过的消息的身份 ID 从而屏蔽掉重复进入的消息，消息转发通道允许采用 TLS 等传输层安全协议实现对密级信息在传输通道的安全保障。

2.4 数据同步与异步处理机制

企业在进行数据交互的过程中不同的业务对于数据处理的时效性的需求也相差很大，因此平台必须能够同时具备同步和异步这两种数据处理的方式。同步处理方式适合那些对及时性的需求非常高的一些情况比如用户的登陆校验，实时库存查询等，在这种情况下，源系统的数据交互请求一旦被触发之后，平台就马上对该请求做出相应的操作并实时地把结果传递到被调用端去，这个过程中对于发起调用的一方来说就是一种同步阻塞的过程。其优点就是非常及时，有很快的响应速度。而异步方式适用于大体量的数据批量同步，不定期的报表打印等问题，这些都不太注重及时性。异步方式中，来源端发送数据之后不需要等待接收方处理完成就可以进行下一步的操作了，平台把数据存到消息中间件中就马上向来源端反馈已接收到，然后在后台进行异步处理。异步处理的方式就是借助了中间的消息队列来做一个缓冲作用，可以实现数据生产和数据消费之间的分离，允许上下级系统各自独立的去处理自己的部分工作，在有大量数据的时候，消息中间件起到了一个缓解压力的作用，不至于让下游系统瞬间压力过大^[2]，平台同时支持定期同步策略，用户可以根据实际需要设置同步任务的间隔时间。针对必须做到数据一致性的跨系统应用需求，平台提供分布式事务最终一致性策略，采用记录事务状态表加补偿的方式，在发生错误的时候能自动补偿或者人为干预使得跨系统的数据最终能够统一起来。

2.5 数据质量控制与异常处理机制

数据质量是评估一个数据交换平台的重要标准，平台基于数据校验、异常监测、纠错及报警等多种方面建立了健全

的数据质量管理机制,在数据接入阶段对获取到的数据进行完整性检验,查看是否存在必填字段为空的情况,数据类型正确与否,数据长度是否超出了设定限制等,针对检验失败的数据将其加入到错误数据队列当中并做好相应的错误日志以便后续的排查及修正工作。数据质量管理融入整个数据集成全流程,在整个数据集成流程中都有相应的数据质量检测措施,可以在第一时间检测出来并且解决所存在的问题,避免次品数据传递到下一个环节当中,保证了数据集成的有效性与正确性。对于异常处理机制上也作了明确的规定,网络层出现链接失败时自动重连,使用指数回退法来决定重连的时间长度;在数据转换时若遇到非正常的格式,则可以设置为默认值替换、跳过这条记录或者停止执行等方法。平台具备完善的数据血缘跟踪的能力,记载每一次数据从源头系统传输至目标系统的所有传送路线,监控告警模块对每一个数据交换作业进行动态监测,一旦指标超过限定界限提示管理员,从而达到数据交换全流程管控的目的。

表 1 数据交换平台异常处理策略分类

异常类型	处理策略	适用场景
网络连接异常	指数退避重试	临时性网络波动
数据格式异常	默认值填充或跳过记录	非关键字段数据问题
业务逻辑冲突	记录详情并触发告警	数据依赖缺失、主键冲突
系统资源超限	熔断降级	CPU/内存占用过高

3 企业信息系统数据交换平台详细设计与实现

在上述关键技术的基础上,此系统的设计与开发采用了分层结构。整个系统从下到上依次为数据源层、采集接入层、消息传递层、数据加工层、数据目的层及管理系统监控体系。数据源层由企业的各种异构的信息系统构成,在此采用适配器的方式接入平台中;采集接入层负责处理各种不同类型的数据源接入逻辑问题,比如有数据库日志监听器、API 轮询调度器、文件扫描器等;消息传递层围绕着消息队列设计而成,依据不同的业务需求分成不同的主题用来对不同种类的数据实行存储以及转发;数据处理层即整个系统的主体业务部分,主要有转换引擎、映射解析器、异常处理器等模块组成。数据目标层用于把经过处理过的数据写入到目的系统中去,也是使用了适配器模式来满足不同目标系统的需求。管理和监控层提供了 web 管理页面,对交换任务进行配置、调度、监控及运维管理等功能操作。对于数据库的设计,平台元数据库包括数据源配置信息、映射规则定义、任务调度配置、执行日志记录、异常数据信息等信息,其中数据源配置表保存各个接入系统之间的连接信息、接入协议类型,映射规则表以 JSON 形式保存源端字段与目的端字段之间一对一或者一

对多的关系,任务调度表保存了交换任务执行周期与优先级等配置信息^[3]。主要编程实现部分,数据抓取模块以模板方法模式来设定抓取过程框架,消息队列包装层提供通用的生产者及消费者接口,转码器是以抽象语法树为基础建立起来的,将用户的自定义映射规则转译成表达式树形式,在执行的时候对输入的数据做分叉遍历计算得到输出的结果。

4 企业信息系统测试与性能评估

为了保证系统的稳定可靠,本文做了大量的测试及性能分析。测试环境安装在国内某公司内网服务器阵列上,安装 3 个节点,每个节点安装的是 8 核 CPU、16G 内存、200GSSD;消息队列池有三台消息队列服务机器,数据源使用普通的数据库,仿真企业的实际业务情况。功能测试包括数据抓取、清洗、传送、同步等功能主要部分设计有 126 个测试案例并且都合格。针对不同结构的数据源接入问题,系统可以实现对各种类型的常用的普通关系型数据库和表达式传送类型的应用程序接口进行采集、入库。数据转换功能实现对字段映射、表达式计算、类型转换等功能的支持,转换成功率为百分百;异常处理功能通过模拟网络中断,数据库链接超时等情况,平台可以正常进行重试或者发送报警等动作;性能测试平台单节点情况下处理单条数据耗时大约为四十毫秒^[4]。并发测试中用性能测试工具开启了一百个并发线程并不断的进行数据交互的操作,平台的吞吐率达到了每秒钟一千二百条记录,cpu 的占用也保持在百分之六十五以下。通过百万测试数据量的一次批量同步测试结果显示:全量同步耗时为 22min,增量同步耗时为 3min,消息队列的削峰能力测试结果是当数据流量突然上升到最大时候的三倍情况下,消息队列可以很好的进行缓存,下游处理系统负载在可承受范围之内。系统的稳定性测试连续不断地运行了 72 个小时,一共处理了大约 5200 万条数据,在此过程中并没有出现消息丢失或者数据重复的情况,系统可用性达到了 99.95%,符合企业的业务需求。

表 2 平台性能测试数据汇总

测试场景	并发数	数据量	平均延迟 (ms)	吞吐量 (TPS)	CPU 占用率
单条处理	1	1000 条	45	22	15%
并发处理	100	10 万条	68	1200	65%
批量同步	50	100 万条	-	757	58%

5 结语

本文对企业信息系统数据交换平台的设计及开发进行系统的探索,从企业的数据分析到非功能性要求出发明确系统建设的目标方向,在核心技术研究上,针对数据采

集与接入、数据转换与映射、数据传输与消息队列、同步异步处理机制以及数据质量控制与异常处理的技术难题进行了详细的论证并做出了相应实现;以关键技术的研究为基础构建了层次化结构体系以及主要的功能组件;经过实测及性能验证其可以有效地解决不同系统之间的信息对接问题,具有较好的高流量、低时延、高效稳定的特性,适用于企业的数据交换场景应用。未来的工作主要围绕着云原生技术的发展,实时计算引擎的应用,以及对数据管理能力的加强来提高平台上人工智能的程度及对企业的价值。

[参考文献]

[1]刘鹏.医院信息系统数据交换平台的设计与实现[D].绵阳:西南科技大学,2017.

阳:西南科技大学,2017.

[2]李云鹏.基于大数据平台的牵引供电管理信息系统的设计与实现[D].南昌:华东交通大学,2020.

[3]董艳.信息系统统一数据访问平台设计与实现[D].武汉:中南民族大学,2016.

[4]徐正雄,刘辉,李思琳,等.阳光采购平台数据自动采集信息系统设计与实现[J].电脑编程技巧与维护,2023(6):106-108.

作者简介:乔敏(1980.11—),毕业院校:电子科技大学,所学专业:计算机科学与技术,当前就职单位:零八一电子集团有限公司。