

基于 SOA 架构的电网数据资产管理平台^①

衡星辰¹, 张诗军¹, 陈 丰¹, 季胜鹏²

¹(中国南方电网有限责任公司, 广州 510623)

²(江苏瑞中数据股份有限公司, 南京 210012)

摘 要: 电网数据已成为企业发展的重要资产, 但电网企业缺乏有效的技术手段来解决电网数据资产的全生命周期监控和管理. 从组件化、松耦合、标准化等架构设计原则考虑, 设计并实现了基于 SOA 架构的电网数据资产管理平台. 设计了一种通过捕获数据库日志方法来实时采集数据资产变更信息, 以及台账管理、全景视图管理、日常监测管理、溯源管理、价值评估管理和流程管理等数据资产管理核心功能, 实现了电网数据资产的标准化、从业务、技术、管理等视角对数据资产进行全方位体现.

关键词: SOA; 电网数据资产; 数据库日志; 数据交易; 资产管理

Grid Data Asset Management Platform Based on SOA

HENG Xing-Chen¹, ZHANG Shi-Jun¹, CHEN Feng¹, JI Sheng-Peng²

¹(China South Grid Corp, Guangzhou 510623, China)

²(China Realtime Database Co., Ltd., Nanjing 210012, China)

Abstract: Power grid data has become an important asset of the enterprise, but power grid enterprises lacks effective technical means to solve the grid data assets life cycle monitoring and management. From the design principle of component, loose coupling and standardization, the power grid data assert management platform based on SOA is designed and realized. It designs a method of capturing the database log to realize real-time acquisition of data assets information change, and account management, management of the panoramic view, daily monitoring and management, traceability management, valuation management and process management data asset management the core function, and realizes the standardized management of grid data assets, and shows the status of data assets from the business, technology, management and other perspectives.

Key words: SOA; power grid data assert; database log; data transaction; asset management

随着信息通信技术的不断进步, 其对中国电力工业的价值贡献正处于量变到质变的关键节点, 而变化的本质就是电力信息通信与电力生产以及企业深度管理的深度融合, 其最终表现形式是电力数据的爆发性增长, 即电力大数据. 电力大数据通过对市场个性化需求和企业自身良性发展的挖掘和满足, 重塑中国电力工业核心价值, 由“以电力生产为中心”向“以客户为中心”转变. 大数据将推动企业向数据资产化和决策智能化发展, 信息部门将从成本中心转向利润中心. 数据渗透各个行业, 渐渐成为企业战略资产. 拥有数据的规模、活性, 以及收集运用数据的能力, 将决定企业

的核心竞争力. 企业战略将从业务驱动转向数据驱动, 智能化决策是企业未来发展的方向^[1].

经过多年的信息化建设, 电网企业已经走过了以流程为中心的阶段, 进入以数据为中心的阶段, 构建数字化企业, 提升企业运营效率是电网发展的迫切要求. 随着电网企业经营业务发展及信息化水平的提高, 亟需高效管理、分析和利用不断产生的各类海量业务数据, 将数据逐步转化为核心资产, 充分挖掘数据资产价值, 将企业管理从“业务驱动”转向“数据驱动”^[2]. 因此, 电网企业逐渐认识到数据是永恒不变的, 是企业始终存在的最有价值的资产.

① 收稿时间:2015-12-31;收到修改稿时间:2016-02-25 [doi:10.15888/j.cnki.csa.005327]

电网企业明确提出要将数据作为公司战略资产进行管理,树立数据资产管理理念,消除专业壁垒,实现企业数据资产利用效益的最大化。运营数据资产的归口管理部门,要加强运营数据资产整合管理和分析利用,挖掘数据资产价值。然后当前的管理手段和机制远远无法满足大数据平台对海量数据资产的管理能力要求,主要包括:

1) 缺乏全局物理视图。如想要知道某一个数据库上面有多少数据表、有多少索引、有多少视图?

2) 缺乏不同维度字典视图及相应 IT 化支持。已存在系统,有哪些关键业务数据,数据资产所有者是谁?谁做过修改或变更?

3) 缺乏问题溯源。某张报表的 A 指标是如何算出来的?原始数据来源是哪里?

4) 缺乏变更管理。为什么这个月数据库里面又多了 100 张表?为什么这张表的一个字段被删掉了,什么时候删掉的?

5) 缺乏数据运行态管理。某张基础表的数据量有多少,数据增长情况如何?

6) 缺乏审计稽核机制等。在设计文档中某一个表有 10 个字段,然后在数据库里面该表为什么有 12 个字段?

企业数据资产管理的主要对象是数据,面对海量业务数据,仅靠传统线下手段,数据台账管理、数据监测、数据模型、数据溯源等资产管理工作将难以开展,无法保证数据资产管理工作效率,因此亟需开发数据资产管理平台,从数据资产核心工作出发,依托支撑平台,实现对电网数据资产的全生命周期的监控和管理^[3]。

1 平台总体架构

面向服务架构(简称 SOA)是一种架构模型,它倡导组件化、松散耦合、隔离关注、标准化等架构设计原则。实际应用可以将其业务功能、信息或流程封装为基本组件-标准化的服务,这些服务通过松耦合的 SOA 架构,为其它应用服务^[4]。SOA 中包含服务提供者,服务请求者和服务代理者,它们分别负责服务发布,服务查找和服务绑定三项操作。服务操作者向服务代理者发布服务,服务请求者通过服务代理者查找所需的服务,并绑定到这些服务上,服务提供者和服务请求者可以交互。对于使用服务的应用,不用关心

服务的提供者是基于什么开发技术、在哪个位置、什么硬件平台提供的服务,这个服务使用的过程完全是松散和透明的。而另一方面,提供服务的业务系统即能得到良好地功能重用,又不会被其它系统侵入,业务系统的技术平台演进也不会影响到使用其服务的业务系统。基于 SOA 架构的电网数据资产管理平台可以在最大程度上减少平台的耦合,为平台整合和扩展提供了理想的解决方案^[5]。本系统总体分成三部分:相关 Web 页面获取模块、Web 信息抽取模块、知识表示模块。系统总体框图如图 1 所示。

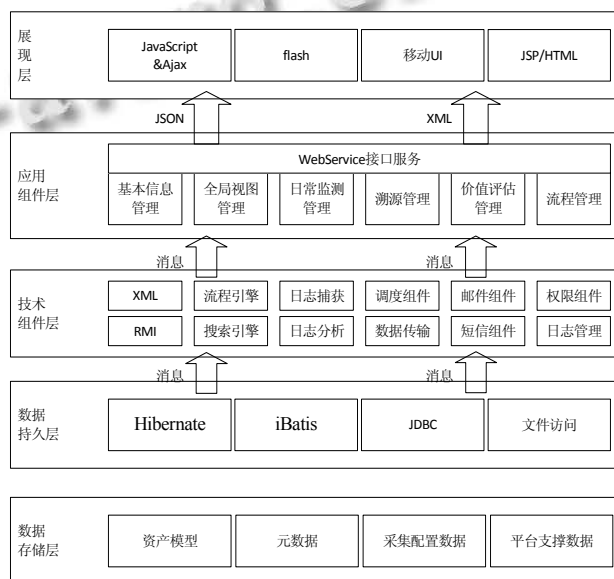


图 1 电网数据资产管理平台总体架构

按 SOA 架构思想,平台总体分为数据存储层、数据持久层、技术组件层、应用服务层、展现层,如图 1 所示。

数据存储:主要是对平台资产模型、元数据、采集配置数据、及平台支撑数据的存储。

数据持久层:采用 Hibernate、iBatis 和 JDBC 相结合的方式,对于量不大且结构比较严谨的数据, Hibernate、iBatis 是一种比较实用的存取方案。对于量大且性能要求高的批量元数据存取,由于 Hibernate、iBatis 性能上无法保障,因此,采用 JDBC 直接读写方式。另外,针对文件类数据,提供单独的文件访问组件。

技术组件层:将平台多种基础服务单独封装、部署,便于上层灵活调用、部署,并基于 Spring Framework,采用 Spring 的事务管理机制以及 Ioc 机制。业务逻辑层为

控制层提供接口调用, 针对分析和采集等应用提供 WebService 接口和 RMI 远程方法调用接口。

应用服务层: 平台提供多种数据资产管理应用服务, 服务通过封装逻辑组件实现, 也可以调用其他业务系统的服务来实现部分功能。这些服务可以被界面层各类应用调用。

展现层: 采用多种 UI 框架, 包括 Flash、Ajax、移动 UI 等, 对于分析类图形展现, 采用 Flash 能够达到很好的交互效果。

2 平台关键技术

电网数据资产管理平台是对企业分布广泛的各种数据资产的实时和有效管理, 因此涉及大量业务系统数据资产信息的自动化采集。在采集方式上要求实时采集并且对原有业务系统影响最小, 而业务系统大部分核心数据都存储在像 Oracle、SqlServer、DB2 等关系数据库中, 因此对该类数据库中的数据交易信息进行实时采集是至关重要的^[6]。本文提出一种通过捕获数据库日志技术来实时采集数据交易信息。该技术的工作原理如图 2 所示。

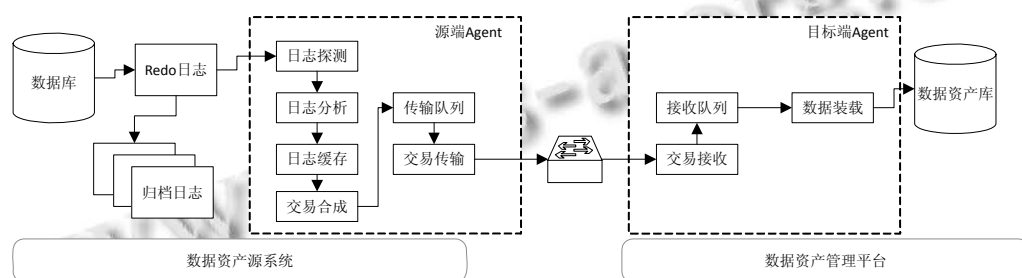


图 2 数据资产自动化采集原理图

在源端和目标端分别安装 Agent 进程, 源端的 Agent 进程对 Oracle、SqlServer、DB2 等日志进行监控, 发现改变及时对目标数据库进行更新。当应用系统在源端向数据库进行任何操作时, 这些信息都将在 RedoLog 中保存, 软件 Agent 通过对实时获取的 Log 日志进行分析, 获得本次操作的交易指令和交易数据, 然后将这些交易指令和交易数据经过格式转化生成紧凑数据格式, 并实时通过网络传送到目标系统。目标系统的软件 Agent 接收数据库包, 经过校验码检查, 确认正确的数据库包后, 调用 Oracle、SqlServer、DB2 等函数按照交易的先后顺序在目标系统中执行该交易。

2.1 日志探测

Agent 通过定期检查 Oracle、SqlServer、DB2 等 Controfile 中记录的当前 SCN 号来判断源端的 Oracle、SqlServer、DB2 等系统是否有新的交易产生。当确认 Controfile 中有新的交易产生时, Agent 获取当前 Redo Log(或者 Transaction Log)组及最新日志文件的位置。Agent 模块根据这些信息将上次抓取时记录的日志位置与本次读取的最新位置之间的 Log 加以分析, 并将这些数据保存在 Log Cache 文件中, 等待下一

步进行交易合成处理。

2.2 日志分析

Oracle、SqlServer、DB2 等数据库的所有更改都记录在日志中, 当我们需要了解数据库中所作的交易时, 可以通过 Agent 中集成的日志分析功能分析 Oracle、SqlServer、DB2 等数据库的日志文件实现。

通过对日志进行分析, 得到该数据库中所有 SQL 指令, 并将这些 SQL 指令生成一种紧凑格式的表达方式。使用紧凑格式能够通过专有转换算法直接转换为 Oracle、SqlServer、DB2 等的内部数据表达格式, 在分析和转载时仅需要进行最小的格式转化即可, 从而提高性能和效率。

2.3 交易合成

通过 Oracle、SqlServer、DB2 等 Redo Log(或者 Transaction Log)分析的交易指令存在以下两个特点:

这些指令是交叉出现的。一个交易(Transaction)中的多条 SQL 指令是非连续存储的; 而多个交易中的 SQL 指令之间是相互穿插的。

Redo log 中记录了所有交易, 包括已经 Commit 的交易和未 Commit 的交易。

为了提高系统的可控制性、保证逻辑完整性及避

免数据丢失,最好将复制的最小单位作为一个交易(Transaction),这样在目标端的交易装载更加容易控制。

同时对于复制的数据而言,在目标系统中只有那些已经 Commit 的数据是有意义的, Rollback 的数据无需复制到目标系统上。因此不是复制每个 SQL 语句,而是对抓取的数据进行交易整合后以交易(Transaction)为单位进行复制,并且只复制 Commit 的交易。

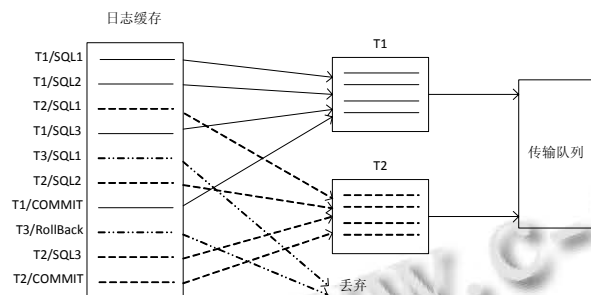


图3 数据库交易合成示意图

如图3所示, Online Log Cache 文件包含 Commit 的交易、没有 Commit 的交易和 Rollback 的交易。交易合成模块首先按照交易序号对 SQL 语句进行划分,然后以交易为单位将已经 Commit 的交易,传递到传输队列,将未 Commit 的交易保存在本地,将 Rollback 的交易作丢弃处理。当交易合成模块通过日志得知保存的未 Commit 的交易已经提交,则该交易将会被立即发送到传输队列中。

2.4 交易传输

为了保证数据传输的安全性和可靠性,交易数据在传输之前首先存入源端 Agent 的传输队列中,传输模块从队列中按照先进先出的原则读取交易数据并将其封装为 TCP/IP 数据包,传送给目标端的接收模块。

在目标端,接收模块在收到传输的交易数据包后,首先根据包头描述的包大小进行传输的合法性检查,判断是否传输完整,如果不完整将记录信息并反馈源端 Agent 重传。如果完整,则将完整的交易数据存入接收队列中,然后由装载模块从队列中严格按照交易的顺序装载交易信息。

2.5 数据装载

数据装载模块将接受队列中的交易数据通过调用 Oracle、SqlServer、DB2 等数据库 I/O 层接口直接将数据通过最底层函数写入数据资产库中,使得装载端的性能和处理能力显著提高。

3 平台功能设计

电网数据资产管理是针对电网企业数据资产的形成、转化、资产运行维护以及资产价值应用的全过程、全方位管理,其目标是针对运营数据资产进行集约、统一、规范管理,开展数据质量的管控与考核,推进数据得到有效的分析和利用,挖掘数据价值,实现数据资产价值最大化。

电网数据资产管理平台基于数据溯源关系及数据关联关系,形成公司数据资产全景视图,采用数据资产实时监测技术,通过源头把控、关键节点监督、防篡改监控等手段,将企业数据逐步转化为可信、可用、在用的数据资产。

基于数据资产管理工作的需求,电网数据资产管理平台包括以下功能。

3.1 数据资产基本信息管理

数据资产基本信息管理主要对于数据资产的分类、台账、编码、业务对象、数据模型、规则和关联关系等进行管理。

1) 据资产分类管理:定义数据资产分类,形成数据资产分类标准,明确数据资产范围。

2) 据资产台账建档:对数据资产建立详细的台账,提供数据资产台账维护功能,可维护数据资产的技术属性、业务属性、管理属性以及基本属性。

3) 一编码维护:对数据资产涉及的编码进行统一管理,定义数据编码规范。

4) 键业务对象管理:对数据资产涉及的关键业务对象进行管理。

5) 据模型管理:建立数据模型,对业务数据项、参考数据等进行统一管理维护。

6) 则管理:对数据资产涉及的质量校验规则、日常监控预警规则以及数据资产价值评估规则等进行管理和维护。

7) 据关联管理:管理数据关联关系,维护数据关联稽核规则,设置跨业务领域的数据间验证关系等。

3.2 数据资产全景视图管理

以数据资产全景视图方式展示全公司范围的数据资产,从业务、技术、管理等视角对数据资产进行全方位体现。

① 全景视图显示:以图形化方式展示数据资产全景视图。

② 全景视图维护:支持全景视图新建、修改、删

除功能, 支持可视化操作, 支持动态配置技术属性、业务属性、管理属性、资产属性等各项属性, 支持关联关系模型定义.

③ 多维度视角展示: 支持从技术视角、业务视角、管理视角、资产视角查看数据全景视图.

④ 数据资产完成率展现: 图形化显示数据资产完成率, 直观反映数据资产完成水平, 为数据资产完成率评价工作提供支撑.

3.3 数据资产日常监测管理

依据数据资产全景视图, 对数据资产设置关键监控点, 日常监测数据资产关键监控点变化情况、数据及时性、完整性、准确性验证、数据篡改情况等.

① 数据资产自动监测: 制定数据自动监测计划, 系统依据各节点评价规则维护内容, 开展自动评价工作, 对超出设定阈值的数据资产进行异动告警.

② 数据资产关键点监控: 基于数据资产溯源图, 在业务及技术层面设置数据资产监控点, 依据监控规则, 对监控点进行自动监测和预警功能.

③ 数据关联监测: 利用跨业务数据间的关联验证规则, 完成跟踪分析.

④ 数据防篡改监控: 对数据库访问制度审查, 制定数据防篡改规则, 依据规则对数据篡改情况进行实时监控.

⑤ 数据资产人工检查: 支持人工检查各节点明细数据, 按照业务定义自动计算结果符合度, 支持现场稽核工作开展.

3.4 数据资产溯源管理

从源头对数据资产进行摸底及梳理, 基于梳理结果管理数据资产溯源全图, 可对数据溯源图进行动态维护以及溯源下钻查询功能.

① 数据溯源节点配置: 对数据溯源节点进行动态定义及配置.

② 数据溯源图可视化展现: 可视化展现数据溯源图, 能支持溯源图向下钻取, 属性查看.

③ 数据溯源图维护: 对数据溯源图的可视化新建、修改、删除等.

④ 溯源图版本管理: 记录数据溯源图的变更历史, 提供版本控制功能.

⑤ 溯源节点数据明细查询: 支持溯源图中各存储节点数据明细查询, 对数据资产人工监测工作提供支撑.

3.5 数据资产价值评估管理

数据资产价值评估模型维护: 提供新增、删除、修改功能, 基于图形化方式对数据资产的价值评估模型进行管理, 关键数据资产及其价值评估模型.

数据资产价值评估: 基于数据资产价值评估模型对数据溯源图各个节点分别进行价值评估, 依据价值评估规则, 对数据可信度、及时性、完备性等进行评价, 并按照价值评估模型, 计算数据资产价值.

3.6 数据资产流程管理

按照数据资产管理办法, 采用信息化手段管理数据资产形成、变更、停用过程, 建立数据资产管理标准化流程, 如图 4 所示.

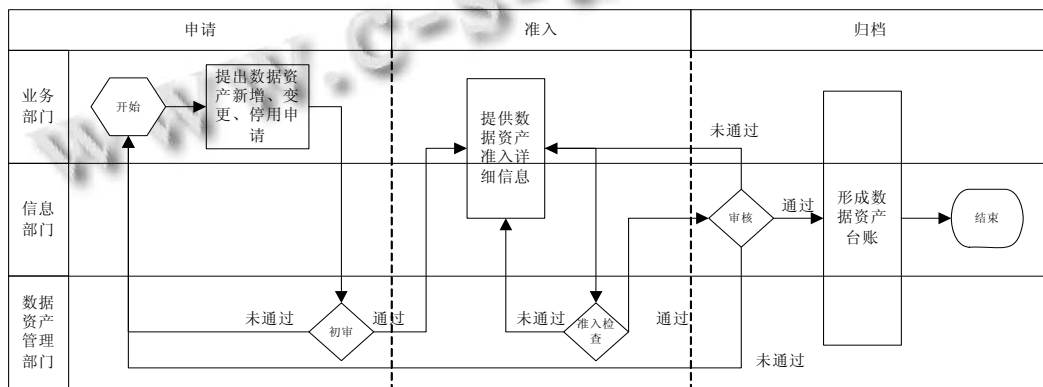


图 4 数据资产管理流程图

① 数据资产新增、变更或停用申请: 提供数据资产新增、调整需求申请功能, 由业务部门提出数据资

产新增、变更或停用申请.

② 数据资产初审: 提供数据资产初审功能, 由数

据资产管理部门对业务部门的数据资产需求进行初审,初审未通过则退回业务部门。

③ 数据资产准入核查:数据资产管理部门对经初步审定后的数据资产进行细致核对,开展准入核查工作,确定数据资产在数据一致性、数据完备性等方面是否符合要求,如有问题,则由业务部门进行数据资产整改。

④ 数据资产整改:业务部门针对数据资产存在的问题进行整改,针对数据资产存在的各类问题,记录数据资产的整改情况,控制数据资产版本变更。

⑤ 数据资产审核:对符合准入规则的数据,由数据资产管理部门负责审核确认。

⑥ 数据资产归档:对审核完毕后的数据资产,业务部门完善数据资产业务属性,信息部门完善技术属性,形成完整的数据资产基础信息,数据资产管理部门负责归档。

4 结语

本文设计实现的电网数据资产管理平台已成功应用南方电网公司,将企业数据作为核心资产管理,通

过对运营数据的全过程有效管理和合理应用,不断挖掘数据价值,实现数据向信息的转化,继而形成业务价值,实现核心业务能力提升、信息价值链扩展和企业运营模式优化。同时,为深化数据治理工作,提升数据的质量,为公司及时、快速的监控企业经营健康情况提供有效途径。

参考文献

- 1 崔吉峰,闫斌,张朋.数据资产管理研究与探索.中国电力企业管理,2014,23(12):94-95.
- 2 李建华,曹路,庄侃沁.华东电网计算统一数据管理平台研究.华东电力,2013,41(1):69-72.
- 3 熊鑫.钢铁企业应用集成主数据管理平台设计与实现.信息技术与信息化,2014,169(1):21-24.
- 4 李萌,魏玮.基于 soa 的主数据管理架构设计及实践.兵工自动化,2015,34(8):49-51.
- 5 董致远.浅谈企业的数据质量管理.中国管理信息化,2015,18(3):78-79.
- 6 屈志毅,王涛,李建旭.DB2 大型数据库容灾备份实时复制系统.计算机系统应用,2012,21(3):198-201.