



基于一体化平台数据中心的数据质量管理在 内蒙古电力公司的应用

张 瑜,潘红芳

(内蒙古电力公司信息通信中心 呼和浩特 010020)

摘 要:在一体化平台数据中心系统建设过程中,为了对数据进行科学有效的管理和质量控制,为决策者提供可信的决策支持,数据中心建设了数据质量管理平台。从信息技术及管理控制两方面入手,对数据质量管理的实现方案进行深入分析,建立了一整套数据质量管理体系,提升了企业数据质量,保证了数据中心系统能够提供准确、科学的数据。

关键词:数据中心系统;数据质量管理平台;数据仓库;数据集市

doi: 10.3969/j.issn.1000-0801.2014.01.022

Application of Data Quality Management Platform in Inner Mongolia Power Company

Zhang Yu, Pan Hongfang

(Information Communications Center, Inner Mongolia Power Co., Ltd., Huhhot 010020, China)

Abstract: The data quality management platform was designed and constructed while the data center was building, which could help to realize scientific and effective data management and quality control. In addition, the trusted data could provide reliable decision support for decision makers. Based on these systems, it was further analyzed that how to build a set of quality management system from two aspects, such as information technology and management control. With the help of reliable data and quality analysis system, the quality of enterprise data was significantly improved, and the accurate scientific data was ensured to be provided by data center systems.

Key words: data center system, data quality management platform, data warehouse, data mart

1 引言

内蒙古电力公司自 2011 年开始逐步建设一体化平台——数据中心项目(以下简称数据中心系统),数据质量管理平台是数据中心系统的关键功能模块。随着数据中心系统价值的不断提升,内蒙古电力集团公司对如何保障数据质量、提升数据资产价值提出了更高的要求。本文以数据中心系统建设中的“数据质量管理平台”为基础,详细地讨论了在数据质量管理方面的具体实施方案及过程。

2 背景及架构

2.1 背景

内蒙古电力公司自 2011 年开始逐步建设数据中心系统,经过几年的发展,已初步建立了基于企业级数据仓库的生产经营决策分析系统,同时满足公司总部及各二级下属单位的决策分析应用要求。通过不断挖掘各个应用系统的数据,进行纵向、横向的立体分析及展现,给各级领导决策层、管理层提供准确、科学的分析依据,为实现公司“一

体化、现代化”的战略目标提供信息化支撑,促进企业全面、协调、可持续发展。

内蒙古电力公司各个业务系统已安全稳定运行很多年,积累了大量的历史数据(其中营销、生产系统的历史数据相对齐全),基本满足企业管理与决策分析的数据需求。但随着系统的不断升级改造,系统使用的时间推移和不断深入,系统难免会产生一些失效数据、无依据来源数据、描述错误或不准确数据以及出现细节数据丢失等现象。

以上现象的存在,导致数据中心在集成各业务系统数据的时候,必须着手解决数据质量问题。只有拥有高质量数据的数据中心,才可以充分挖掘数据潜在的价值,进而为公司带来效益。

2.2 建设意义

数据是企业无形的资产,要让数据最终产生价值,数据的准确性是十分重要的,如果数据质量的问题得不到及时解决,将直接影响企业管理者的科学决策,所以,建设一套完善的数据质量管理体系就显得尤为迫切与重要。如图1所示,数据质量的提升过程是一个不断增值的过程,由于领导决策需求,产生了数据质量需求,进而需要各业务系统改造数据、提升数据质量,才能提供更有效的决策支持,以此循环,使数据不断创造价值。

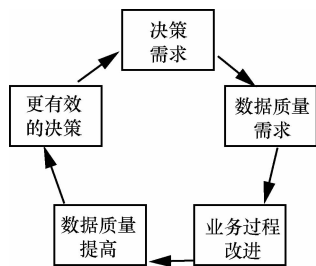


图1 数据质量管理平台的建设意义

2.3 建设原则

(1)健全机制、分级负责

公司成立“数据质量管控小组”,全面负责数据质量平台的建设和实施工作,保证数据质量管理平台在实施过程中及时、有效地沟通,从而更好地推进实施管理工作。

(2)分步实施、稳步推广

按照分步实施、稳步推广的原则开展数据质量管理平台的实施工作,有利于有效控制实施风险,逐步扩大系统的应用范围,保证数据的准确、安全,提高系统价值。

(3)巩固成果、持续优化

数据质量管理平台的建设,始终伴随着管理思想的变革和业务系统的数据改造。在实施过程中,将通过改造业务数据,梳理业务流程,提升管理水平。

3 关键技术及功能架构

3.1 技术架构

数据中心通过采用数据采集工具,将各业务系统数据抽取至数据仓库之中,并以此数据为基础,建立数据质量管理平台,整体技术架构如图2所示。

3.2 数据来源

从图2可以看到,数据质量管理平台的数据源包括以下几部分。

- 内蒙古电力公司的各业务系统,包括生产 PMS 数据、供电可靠性数据、营销 MIS 数据、营销 95598 数据、营销电能量采集数据、人力资源数据。
- 数据仓库中的数据,包括 ODS(操作型数据存储)数据、EDW 数据(明细数据)、CDW 数据(汇总数据)、DM 数据(数据集市数据)。

3.3 疑问数据衡量标准

数据质量可以从数据的完整性、唯一性、准确性、规范性、一致性、关联性、及时性 7 个方面来衡量。

(1)数据的完整性

数据的完整性用于度量哪些数据丢失了或者哪些数据不可用,如果数据库中存储有不正确的数据值,则该数据库称为已丧失数据完整性。

(2)数据的唯一性

数据的唯一性用于度量哪些数据是重复数据或者数据的哪些属性是重复的,包括以下 3 种情况:同一类数据的来源是唯一的;主外键的约束性;对数据采用同一种处理逻辑。

(3)数据的准确性

数据的准确性用于度量哪些数据和信息是不正确的或者数据是超期的,主要反映了数据的真实度,如系统原用户数是 100,但由于数据统计不准确,结果是 101。

(4)数据的规范性

数据的规范性用于度量哪些数据未按统一格式存储,包括数据来源是否合理和数据存储类型是否规范,如某电量数据应该按照数值型存储,但由于数据存储不规范,按照字符型存储。

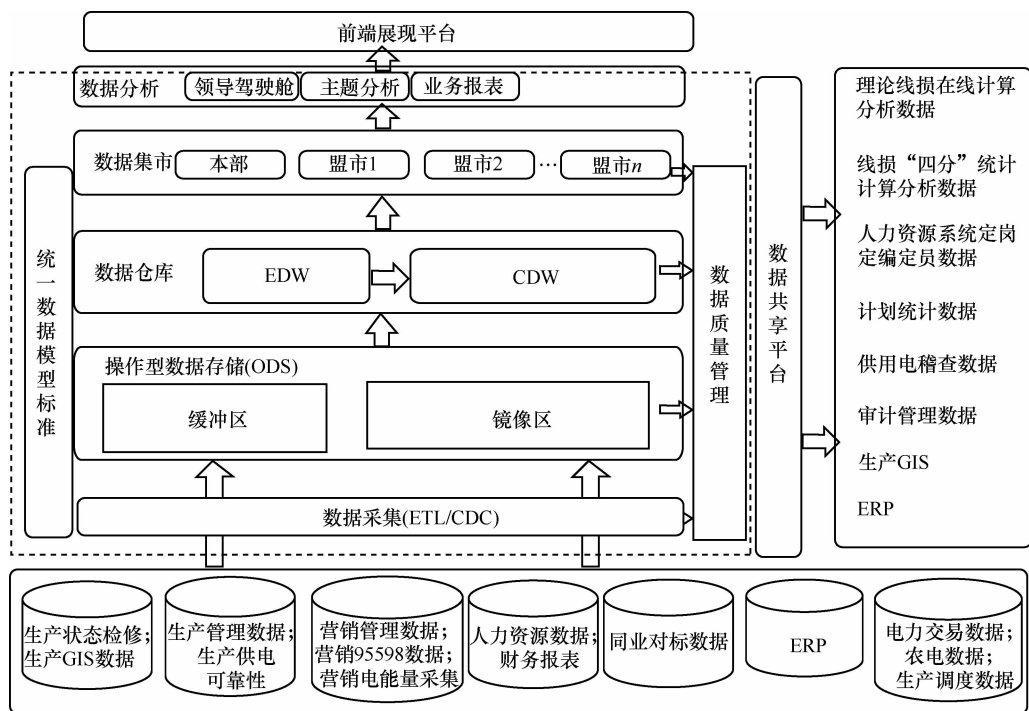


图2 数据质量管理平台的技术架构

(5)数据的一致性

数据的一致性用于度量哪些数据的值在信息含义上是冲突的,主要表现在数据口径是否一致,如电费信息,按照营销口径统计的结果与按照财务口径统计的结果不一致,在统计数据时必须区分。

(6)数据的关联性

数据的关联性用于度量哪些关联的数据缺失或者未建立索引。一般情况下,在源系统中存在关联关系的对象,在目标系统中依然会存在,否则,数据会存在关联性异常。

(7)数据的及时性

数据的及时性用于衡量数据是否按照规定的时间进行抽取、清洗、处理、加载、展示。如果在某个环节出现延迟或者异常,会直接影响后续环节,并导致最终结果数据的不可靠。

3.4 疑问数据过滤规则

在数据质量管理平台建设初期,根据各业务部门工作人员提出的疑问数据判断标准,结合实际数据的存储形式及现状,数据中心项目组拟定了疑问数据过滤规则,所有疑问数据均在该规则下产生。随着系统不断的深化应用及疑问数据种类的增加,该过滤规则会不断扩充,不断完善。

3.5 实现方式

数据质量管理平台作为数据中心系统的一项关键功能模块,采用的技术实现方式与数据中心系统一致,即采用ETL抽取工具,按照已经设置好的规则,将疑问数据加载到数据质量管理平台,再通过页面展示工具进行分析展现。

3.6 功能架构

数据质量管理平台从生产PMS数据质量管控、生产可靠性数据质量管控、人事数据质量管控、营销MIS数据质量管控、营销电能采集数据质量管控、营销95598数据质量管控等方面入手,辅助内蒙古电力公司各业务系统提高自身数据质量,功能架构如图3所示。

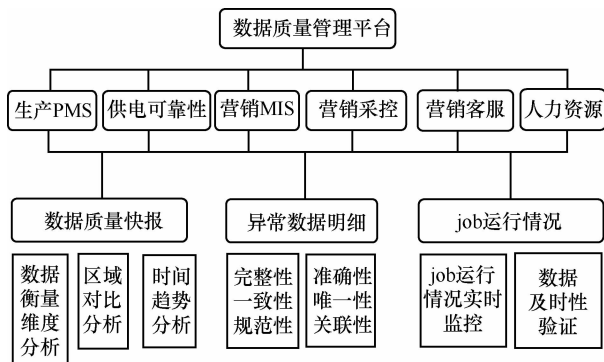


图3 数据质量管理平台的功能架构



检索			
问题类别:完整性▼	所属业务:生产▼	查询	
当前显示: 第1至第5条(共5条) 第1页/第1页			第一页上一页下一页最后页第 1 ▼ 页
问题类别	库表	所属业务	备注
完整性	WZL_MC_WZXX_1000	生产	输电设备完整性问题数据
完整性	WZL_MC_WZXX_1001	生产	变电设备完整性问题数据
完整性	WZL_MC_WZXX_1002	生产	两票管理完整性问题数据
完整性	WZL_MC_WZXX_1003	生产	运行检修完整性问题数据
完整性	WZL_MC_WZXX_1004	生产	缺陷管理完整性问题数据

图5 数据质量管理平台——疑问数据分类分析

度进行汇总展现:

- 页面顶部是条件筛选,包含“问题类别”及“所属业务”;
- 页面下方按照问题类别(数据质量衡量标准)进行分类统计,列出疑问数据所属业务、具体类别、所对应的库表等信息;
- 点击某一行记录,可以查看详细的疑问数据。

图6是数据质量管理平台——job运行情况监控的页面截图。

使用者:数据中心系统运维人员。

job运行情况监控是为满足数据及时性要求而开发的监控页面,该图是对job(抽取数据的程序)运行情况进行分类统计,并查询错误日志;对job抽取策略按不同主题进行详细描述,包括成功数量、失败数量、正在执行数量、执行成功率、执行时长。

同样,如果发现异常,可以进一步查看明细信息,进行追溯、分析。

4 建设成果

数据质量管理平台自规划建设以来,受到了内蒙古电力公司各部门的大力支持,平台应用情况良好。使用者包括公司生产技术部、营销部、财务部、人力资源部等核心部门及10家二级单位;覆盖业务系统包括生产PMS、生产供电可靠性系统、人力资源系统、营销MIS、营销电能采集系统、营销95598系统六大系统;出具数据质量快报30个;提出292项疑问数据,共计71 854条记录,为辅助提高各业务系统数据质量做出了积极的贡献。

5 未来展望

展望未来,数据质量管理平台还需要不断深化应用,未来的数据质量管理平台要从完善疑问数据过滤规则及

加强数据质量改进制度这两个角度进一步建设,最大限度地消除疑问数据,提高数据质量,辅助领导做出科学的决策,帮助管理人员提升管理水平,最终完成数据→知识→信息→价值的转换,体现数据中心的价值,为公司带来经济效益和社会效益。

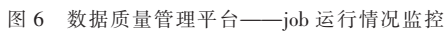
5.1 持续不断地完善疑问数据过滤规则

为了持续不断地完善疑问数据过滤规则,还需要做好以下几件事。

- 定期与业务部门沟通,完善过滤规则:由于业务部门的业务规则可能发生变化或者开发人员发现新的疑问数据,所以,需要定期与业务部门进行沟通,清除已经不适用的过滤规则,补充新增的过滤规则。
- 完善疑问数据检查方式:主要分为全部检查和抽样检查。全部检查是每天对所有的疑问数据过滤规则进行检查;抽样检查是不定期地对业务系统数据进行查询、分析,以便发现新的疑问数据,完善过滤规则。
- 完善业务系统数据库的更新流程:因为业务系统每次更新数据库,就有可能新增数据表或者更改表结构,要想全方位地提升数据质量,必须对新增的数据进行持续监控。所以,需要采用合理的管理手段,完善业务系统数据的更新流程,即在业务系统完成数据库的更新之后,系统厂商及时通知数据中心项目组,从而及时分析新增数据的数据质量。

5.2 加强数据质量改进制度和监督制度机制的建立

加强数据质量改进制度和监督制度机制的建立,需要从提高员工对数据质量的重视程度、完善管理制度、采用先进技术3个方面入手,持续跟进数据质量的总体水平,把数据质量指标纳入考核,建立完善的数据负责制度,确



147



等方面的防护,另外在应用层防火墙上配置服务器隐藏、Web 隐藏将服务器区域的信息隐藏起来,保障内部应用的安全。

(1)业务系统采用 SSL 加密访问

SSL 加密是一个安全通信通道,它可以在用户和服务之前建立一个安全的加密通道,用户通过加密方式访问业务数据,最大程度地降低系统遭受会话监听和劫持攻击的风险。要实现业务系统的 SSL 加密访问,可以在服务器区域前部署 SSL 加密设备,也可以将业务系统结构改成 HTTPS 形式。

(2)强化口令策略

增加口令强度自检模块,检测系统用户口令长度、强度等是否符合强口令要求。如果修改的口令不符合口令策略要求时,系统阻止用户登录,并提示用户重新更改密码,直到口令强度符合要求。系统登录时设置登录次数限制,防止遭受暴力攻击风险。

(3)网络应用隔离

应用系统数据需要与互联网隔离开,以防止信息泄密出去,保证内部数据的安全。通过建设虚拟桌面的方式将整个物理网络隔离成两个不同的网络,用户访问互联网的数据在本地网络中,用户访问业务系统数据在加密的专网中,并且通过策略控制,禁止在虚拟桌面中使用外设(USB、打印机、COM 等),禁止虚拟桌面与本地桌面通信,实现两个网络的隔离。

(4)访问审计

审计是保障信息系统安全的重要手段,而且应用系统的跟踪、监控及灾难恢复都依赖于日志系统。通过部署数据库审计系统和行为审计系统可以有效地记录访问应用

的行为和数据。

(5)数据备份及恢复

为了防止人为的失误或破坏,应在异地建设灾备中心,一旦主数据中心出现故障,异地数据中心可以正常工作,保障业务不中断,数据不丢失。

4 结束语

随着信息安全形势的日益严峻,对信息安全的高度重视,业务应用系统的建设和使用已经不能仅停留在系统可用性功能上。应用系统的设计、开发、运行都必须确保网络与信息的安全。本文以一个发电公司——金堂公司的信息安全为例,在现有的网络和信息化建设情况下,开展了详细的分析与安全设计。本文对发电企业业务系统的安全加固建设、双网隔离、用户管理、移动接入安全有一定的参考价值。

参考文献

- 1 国家电力监管委员会 5 号令. 电力二次系统安全防护规定, 2004
- 2 国家电力监管委员会电监安全 34 号文件. 电力二次系统安全防护总体方案, 2006

[作者简介]



余颖,女,国电成都金堂发电有限公司高级工程师、信息中心主管,主要研究方向为信息化安全及系统建设。

(收稿日期:2013-08-01)

(上接第 147 页)

参考文献

- 1 侯延湘. 数据质量管理及在数据仓库实现探讨. 吉林大学硕士学位论文, 2007
- 2 刘涛. 青海移动数据仓库数据质量控制方案设计与实现. 北京邮电大学硕士学位论文, 2011
- 3 付威. 金融行业数据仓库项目的数据库质量管理应用. 北京邮电大学硕士学位论文, 2006
- 4 杨凡涛. 关于省级数据集中后税收数据质量管理与应用研究. 中国海洋大学硕士学位论文, 2009
- 5 宋敏, 覃正. 国外数据质量管理研究综述. 情报杂志, 2007(2)

[作者简介]



张瑜,女,内蒙古电力公司信息通信中心工程师,主要从事电力信息化网络及数据建设、管理和规划方面的工作。

潘红芳,女,内蒙古电力公司信息通信中心工程师,主要从事电力信息化建设、发展和规划方面的工作。

(收稿日期:2013-08-01)