



内蒙古电力数据仓库的应用

张瑜,潘红芳

(内蒙古电力(集团)有限责任公司信息通信中心,内蒙古 呼和浩特 010020)

摘要:随着电力信息化的发展,电力数据的积累速度越来越快,如何使大数据产生价值、为公司科学发展提供支持的议题被越来越多地提及。内蒙古电力公司从 2011 年开始建设一体化平台数据中心,当前已经完成了公司各业务口的数据集成工作,为数据利用提供了基础。依据建设的一体化生产经营决策分析系统建设情况,从应用需求、应用技术、数据分析挖掘 3 个方面介绍了内蒙古电力公司在大数据背景下的数据仓库应用的探索与尝试。总结了建设成果与建设经验,得出做好分析展现应用的关键是需求管理,核心技术是数据挖掘的结论。

关键词:电力大数据;数据中心;数据仓库

中图分类号:TP311.5

文献标识码:A

doi: 10.11959/j.issn.1000-0801.2016069

Application of Inner Mongolia electric power data warehouse

ZHANG Yu, PAN Hongfang

Inner Mongolia Power Information Communications Center, Hohhot 010020, China

Abstract: With the development of electric power information, the accumulation of power data is more and more fast, and the issue of how to make big data to support the scientific development of the company is increasingly mentioned. In the year of 2011, Inner Mongolia Power Company started to build integrated platform data center. The company has completed the company's business data integration work, and provided a basis for the use of data. Based on the construction of integrated production and management decision-making analysis system, the Inner Mongolia Power Company in the data warehouse application exploration and attempt was introduced from the application needs, application technology, data analysis and mining. Summary of the results and experience of construction shows that the key of application analysis is demand management and the technology is datamining.

Key words: electric power big data, data center, data warehouse

1 引言

随着近年来信息化发展速度的加快,互联网、信息系统、物联网、ERP(enterprise resource planning,企业资源计划)等技术手段纷纷被引入电力行业,给电力行业注入了新的活力。行业内许多公司逐年建设并完善了用电营销、生产管理、财务管控、ERP 等近 20 个业务系

统,提高了各部门的工作效率,提升了部门管理水平,也产生了大量的数据。接着面临的问题就是如何应用数据并挖掘数据的价值,使数据能够服务于企业发展战略。为了实现数据即价值的战略目标,内蒙古电力(集团)有限责任公司(以下简称内蒙古电力公司)在 2011 年开展了一体化平台数据中心的建设工作,当前已完成了 3 期建设,集成了电力公司的核心数据。电力行业数据存在



种类多、数据量大、信息量大、关乎国计民生的特点,针对这些特点,电力数据的应用也分为外部应用和内部应用两部分。外部应用是透明的,和社会其他行业或个体是关联的,内部应用是服务于公司发展、应用于电力内部网络的。本文在电力大数据的背景下,以数据中心系统建设中的生产经营决策分析系统为基础,介绍了电力数据的内部应用。

2 应用背景及建设意义

2.1 背景介绍

随着内蒙古电力公司信息化建设的不断深入,各主营业务系统的建设相继完成,公司管理层、决策层渴望能够及时、准确地掌控公司财务状况、生产状况、设备情况、人员结构等,以帮助提高企业的整体效率和执行水平,减少执行成本,提高企业的资产利用率,实现企业整体盈利水平的改善和提升。经过多年的信息化建设,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。公司已经形成了大数据的应用基础,满足了进一步挖掘数据利用价值的条件。

2.2 建设意义

随着公司的不断发展,公司领导决策层需要基于各业务系统大量的数据进行决策分析,为此,必须对各业务系统的数据进行关联分析和钻取挖掘,进而快速精准地为领导决策层提供有力依据。因为业务系统数据分散,公司领导只能通过周报、月报、季报的纸质文档,查看公司整体运营的情况,数据的及时性、完整性无法得到保证。可以在数据仓库的基础上,面向不同用户需求全面建设生产经营决策分析系统,以解决跨专业数据分析问题,使数据能够服务于决策分析。

3 应用技术及功能架构

3.1 技术架构

数据中心技术架构如图1所示,各层介绍如下。

(1) 前端展现层

前端展现平台是指门户(portal)层面的集成,通过与应用门户进行集成,实现数据中心主题分析、业务报表、领导驾驶舱等功能的对外集中展现。通过采用商务智能(business intelligence, BI)软件,提供丰富的图表展现、数据

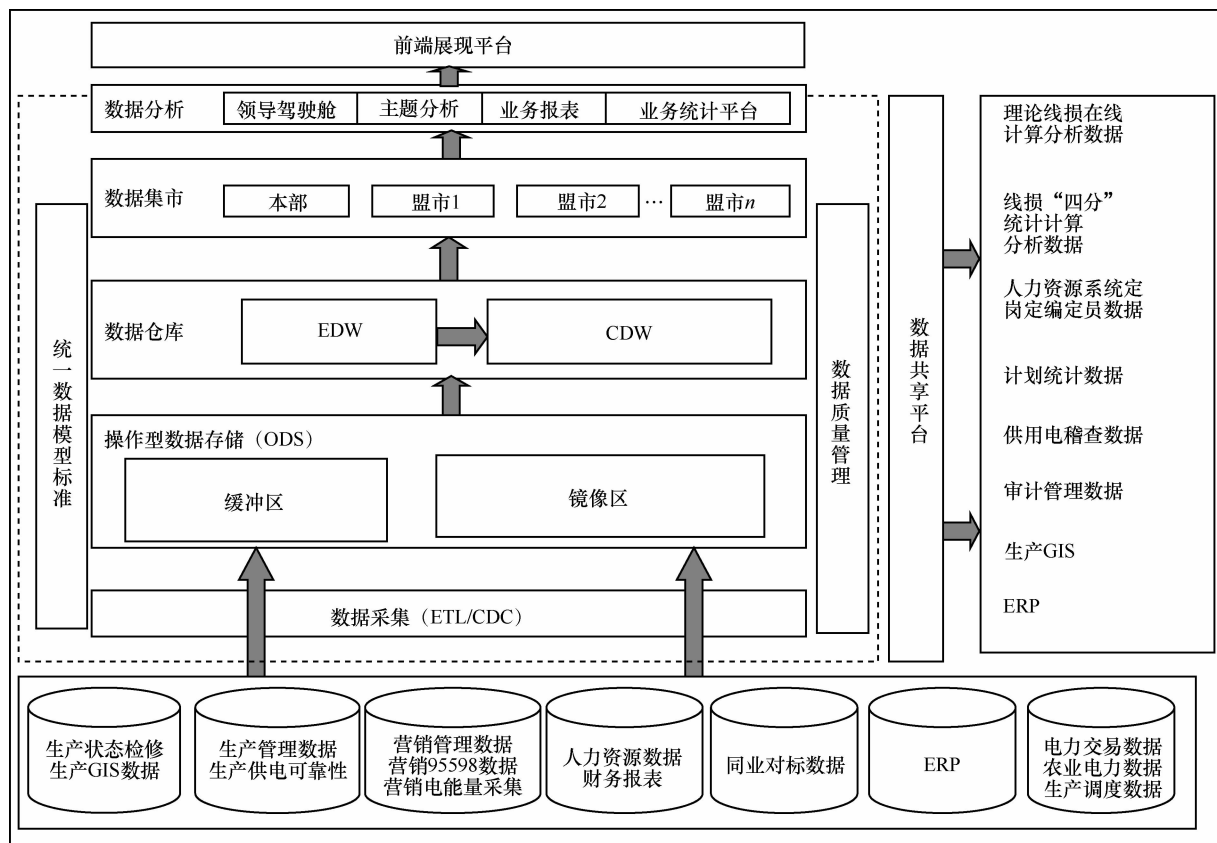


图1 数据中心技术架构

下钻上钻、直观易用的操作。全面支持“公司总部决策层”、“公司总部管理层”、“公司总部业务层”、“二级单位决策层”、“二级单位管理层”、“二级单位业务层”的决策、分析和查询应用。

(2) 数据集成层(数据集市、数据仓库、操作型数据存储)

数据的集成是指在数据源层次上的集成机制,通过 ETL (extract-transform-load, 抽取—转换—加载)/CDC (change data capture, 改变数据获取)等技术手段,实现与现有应用系统的数据共享,提高数据的利用率。

从数据资源存储来看,数据中心划分了 ODS(operational data store)、EDW(enterprise data warehouse)、CDW(corporate data warehouse)、DM(data mart)、DS(data share, 数据共享)五大区域,将数据中心抽取到的各业务系统数据进行合理规范的存储。

其中,ODS 是缓冲区(操作性数据存储区),存储从业务系统获取的原始数据,减轻了对业务系统服务器的压力,满足了实时性需求;EDW 是明细数据存储区,可对数据进行历史追溯;CDW 是汇总数据存储区,只包含汇总数据,提高了数据利用效率;DM 是数据集市区,是为满足个性化需求而提出的存储区;DS 是共享数据区,统一存储各部门共享需求的数据,提高了数据仓库的安全性。

在合理划分存储区域后,为了统一规范数据内容,使

数据能够最大限度地得到利用,借鉴国家标准及行业标准,并结合业务系统数据现状,建立统一模型及统一编码,形成了内蒙古电力公司的统一编码体系。

(3) 数据抽取层

数据抽取层运用多种技术手段将源业务系统数据抽取至数据仓库,数据中心目前已经接入人力资源系统、PMS(power production management system, 工程生产管理系统)、营销 MIS(management information system, 管理信息系统)、财务报表、同业对标、生产输变电可靠性、生产供电可靠性、营销 95598 呼叫中心、营销电能采集等业务系统数据。

为集成数据,数据中心建立数据采集平台后,通过采用数据实时复制工具及 ETL 数据抽取工具,将各业务系统数据抽取至数据中心,同时考虑到有些数据没有进入系统而是以文档的形式存在,为收集文档数据,建立了业务统计平台。

(4) 数据共享层

数据共享层通过数据仓库共享区为各业务部门提供数据共享服务,目前已为生产技术部、发展策划部、企业管理部、审计部、人事部、办公室、财务部、稽查局等部门和单位提供了共享数据服务,满足了各业务部门的数据需求,提高了部门间的共享效率。

3.2 功能架构

数据中心功能架构如图 2 所示,各部分的主要功能介

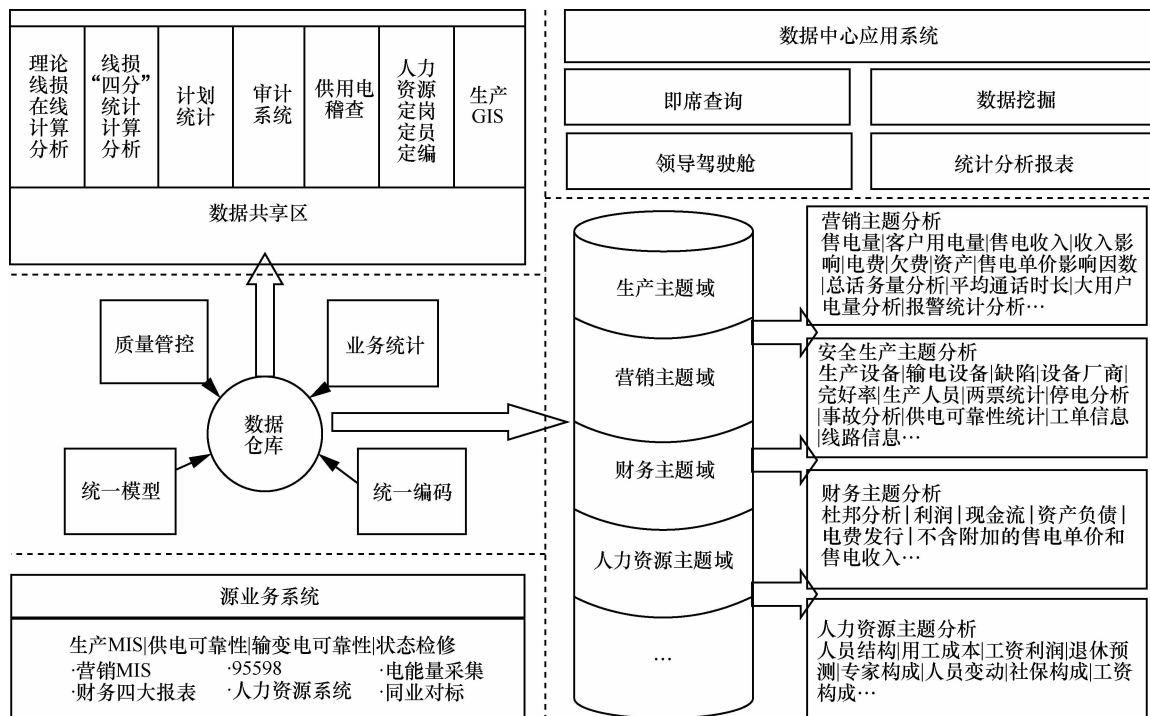


图 2 数据中心功能架构



绍如下。

(1) 源业务系统

以公司建设的各业务系统作为数据中心的数据源。

(2) 数据仓库

依照统一的模型和编码,将数据集成到数据仓库。以业务统计来保证数据的完整性,以质量管控来保证数据的质量。

(3) 数据共享区

提供统一的共享平台,为各业务系统提供数据服务。

(4) 数据中心应用系统

通过数据中心应用系统,实现以集成各业务关键指标为领导决策支持核心的领导驾驶舱功能;以横向协同生产、经营、财务、人力资源关键业务指标为部门业务交互中心的财务报表和同业对标主题分析功能;以集成本部门主要生产经营指标为部门辅助决策入口的管理桌面功能;针对具体指标的面向公司管理层的生产经营指标分析功能。整个展现分析平台在纵向上各级应用的层次分明,在横向上对不同业务指标间的关联分析紧密,可在纵向上对业务指标进行深入挖掘。

3.3 应用技术

3.3.1 商业智能(BI)分析

决策分析平台,通过采用商业智能(BI)软件,开发各种页面,以饼图、柱图、趋势图、面板、表格等方式,综合展现生产、营销、人力资源等核心业务系统数据,辅助决策管理人员做出科学决策。在具体实现决策分析方面包括如下内容:

- 搭建面向本部、盟市的数据集市,满足不同人员的决策数据需求;
- 建设公司领导驾驶舱,集成公司营销、生产、人力资源、财务等业务部门数据,满足公司决策层需求;
- 建设盟市领导驾驶舱,集成本单位营销、生产、人力资源、财务等业务部门数据,满足二级单位决策层需求;
- 建设面向管理层的业务主题首页,集成本业务口关键数据;
- 建设面向业务人员的单指标分析页面,对单项指标进行多维度分析。

通常在前端展现的时候,最终用户的类别由于其业务级别和工作的不同而有所不同。

在营销综合分析页面中,公司领导不仅可以看到关于

售电量的滚动信息,并且能够直观获得售电区域、构成比例、时间趋势,本期、同期等不同维度的售电量信息,通过数据层层钻取,最终看到最细粒度的明细数据。

3.3.2 统一数据模型和编码

统一数据模型和编码是确保数据的准确性、一致性、完整性的手段,需要建立在对业务系统的全面掌握、对未来发展趋势的合理分析基础上。为实现预期目标,需要对业务的深度分析,对标准的灵活借鉴,同时还要结合实际进行个性化设置,实现步骤如下所述。

(1)了解业务系统业务现状、编码现状,分析设计规范业务数据存储模型。

(2)构建统一编码和模型,规范编码信息的组织和表达。

(3)在 ETL 过程中对数据进行编码对照转换,数据清洗和汇总,按照统一模型进行存储。

(4)按照统一编码和统一格式进行共享、展示等应用。

3.3.3 数据存储整合技术

通过对软硬件的整体优化,比传统数据库服务器运行高效。例如在电力行业,率先采用 Oracle 一体机作为数据仓库服务器,比传统数据库服务运行高效、稳定,一体机响应时间比原有数据库环境最高提高 17 倍。

运行期间系统实现复杂的逻辑运算,包括去除空值、类型转换、数据拆分与聚合、数据行转列等,利用任务调度,保证逻辑运算的前后级关系,最终保证数据中心能够提供准确、规范的数据服务。

将 Oracle 物化视图应用在数据仓库中,既减小了对源端的服务器压力,又提高了数据抽取效率,保证了数据的时效性和准确性。

4 建设成果

基于数据仓库,数据中心建设了跨专业的生产经营决策分析系统,以提供丰富的图表展现方式,全面支持决策层、管理层、业务层的决策、分析、查询和应用。面向决策层,建设了领导人管理驾驶舱,以集中展现公司各个业务口的关键指标,辅助查看公司整体运营情况,解决了以往因为业务系统数据分散,公司领导只能通过周报、月报、季报的纸质文档查看公司整体运营情况的问题。面向管理层,建设了管理桌面,以集中展现本业务口的主要指标,部门领导可以通过该页面方便地查看本部门的运行情况。面向业务处理人员,数据中心建设了多维度指标关联分析、

预测分析,自动形成的明细报表页面,为业务人员的日常工作提供信息化支持。用户的专业不同,所关注的要点也不同,基于此,为每个专业的用户制定了对口的数据分析主题域,根据业务的不同,当前建设的主题域包括生产主题、财务主题、人力资源主题、营销主题、调度主题、电力交易主题等。

5 未来展望

无论是对于数据仓库的建设还是大数据的应用,业内比较认可的划分为5个阶段,如图3所示。

第1阶段:数据仓库的基础建设,主要实现数据的集成。在此基础上进行一些基础分析应用,从而了解“以前发生了什么”。

第2阶段:在有了各类数据并且积攒了一定时期的历史数据后,通过关联数据和历史趋势来分析“为什么会发生”。

第3阶段:应用更多的基于数据挖掘的预测应用,通过建立预测模型进行趋势分析,预测出未来的发展趋势,分析“将来会发生什么”。

第4阶段:在前3个阶段发展的基础上,基于大量的历史和实时数据,对预测趋势和希望目标进行比较,进而分析出改进方案,达到“希望发生什么”的效果。

第5阶段:将数据仓库和运营系统进行有机整合,实时反映公司的运行状况,并随时根据变化进行分析反馈,达到“实时监控”的效果。

当前,公司的生产 GIS、配电系统正在建设中,电能量采集终端还没有覆盖全网,地调系统还没有实现大集中。

数据中心还缺乏这些核心数据,对于全网整体的分析还缺少数据支撑。数据中心还处于第一阶段与第二阶段之间,随着系统建设的逐步完善,数据中心未来会建设非结构化数据平台、海量/准实时数据平台、地理信息平台。在这些数据平台的基础上,建设预警平台、报表平台、异常分析平台、对标平台等,逐步达到预测、指导、监控的效果。

6 结束语

基于一体化平台生产经营决策分析系统的建设,总结出做好分析展现应用的关键是立足公司全局(面向决策层建设的领导驾驶舱)、精选关键指标(面向管理层建设的管理桌面)、紧抓业务需求(面向业务层建设的单指标分析与报表统计)。数据量大只是大数据概念的一部分,大数据技术的核心应该是数据的挖掘应用技术,在数据集成的基础上做分析展现是大数据应用的重要一环。未来的电力大数据时代应该是结构化数据、非结构化数据和地理信息等有效融合的时代,利用数据提升决策管理水平、改进工作流程和工作方式、优化资源利用,使数据持续产生更大价值。

参考文献:

- [1] 涂子沛. 大数据:正在到来的数据革命[M]. 桂林:广西师范大学出版社, 2013.
TU Z P. The big data revolution[M]. Guilin: Guangxi Normal University Press, 2013.
- [2] 赵刚. 大数据技术与应用实践指南[M]. 北京:电子工业出版社, 2013.

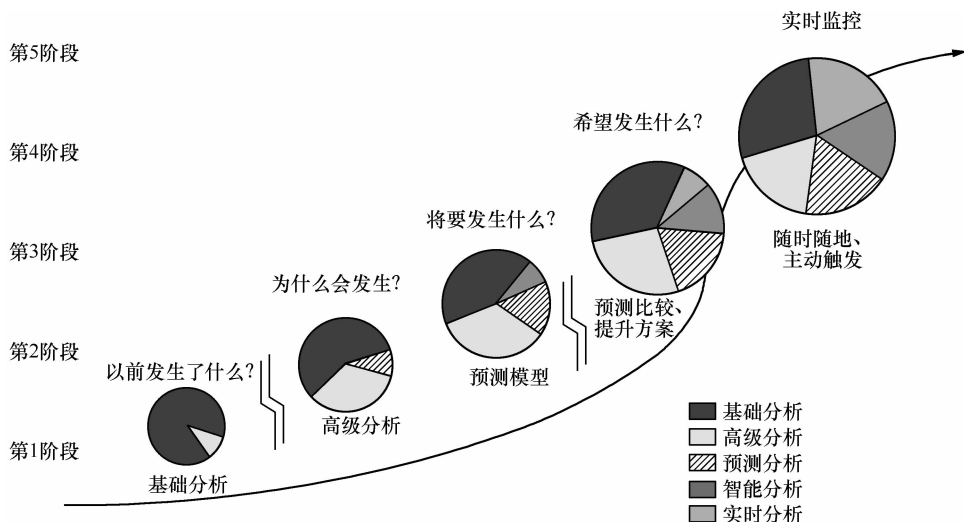


图3 大数据应用的5个阶段



ZHAO G. Big data technology and application practice [M]. Beijing: Publishing House of Electronics Industry, 2013.

- [3] 李建中, 刘显敏. 大数据的一个重要方面: 数据可用性[J]. 计算机研究与发展, 2013, 50(6): 1147-1162.

LI J Z, LIU X M. An important aspect of big data: data usability[J]. Journal of Computer Research and Development, 2013, 50(6): 1147-1162.

- [4] 孙柏林. “大数据”技术及其在电力行业中的应用[J]. 电器时代, 2013(8): 18-23.

SUN B L. “Big data” technology and its application in electric power industry[J]. Electrical Age, 2013(8): 18-23.

- [5] 王继业. 大数据与电力企业[J]. 电力信息与通信技术, 2012, 10(8): 7.

WANG J Y. Big data and electric power enterprises[J]. Electric Power Information Technology, 2012, 10(8): 7.

- [6] 中国电机工程学会电力信息化专业委员会. 中国电力大数据发展白皮书[R]. 2013.

Electric Information Professional Committee of China Electrical

Engineering Society. China electric power big data development white paper[R]. 2013.

[作者简介]



张瑜(1980-),女,内蒙古电力(集团)有限责任公司信息通信中心高级工程师,主要从事电力信息化建设、管理、规划和系统运维等方面的工作。



潘红芳(1978-),女,内蒙古电力(集团)有限责任公司信息通信中心高级工程师,主要从事电力信息化建设、发展和规划方面的工作。