

大型企业 全业务统一数据中心 技术及应用

Technology and Application of Full-Service Unified Data Center
for Large Businesses

吴文宣 主编



中国电力出版社
CHINA ELECTRIC POWER PRESS

大型企业 全业务统一数据中心 技术及应用

Technology and Application of Full-Service Unified Data Center
for Large Businesses

吴文宣 主编



中国电力出版社
CHINA ELECTRIC POWER PRESS

内 容 提 要

电力数据的使用贯穿于电力产业的各个环节,依托全业务统一数据中心对数据价值进行深度挖掘,实现“数据转化资产”“数据转化智慧”以及“数据转化价值”,以新兴技术驱动企业创新化、智能化,助力电网迈进全景实时的电网时代已经到来。

国家电网公司在该领域开展了相关技术研究与应用实践工作,并取得了一定的成果。本书结合国家电网公司全业务数据中心研究成果,着重从数据中心现状及发展趋势,以及全业务统一数据中心概述、总体设计、关键技术、建设与管理、案例与实践等方面入手,以技术结合实例的方式全方位、多视角地展现全业务数据中心给传统电力行业带来的发展创新和变革,为电力行业向能源互联网转型,重构企业价值链,增强核心竞争力提供了重要的参考依据。

本书能够帮助读者了解电力行业全业务数据中心的发展现状,给电力工作者和从事其他行业数据中心相关工作的研究人员和技术人员在工作中带来新的启发与认识。

图书在版编目(CIP)数据

大型企业全业务统一数据中心技术及应用/吴文宣主编. —北京:中国电力出版社,2018.8
ISBN 978-7-5198-2149-4

I. ①大… II. ①吴… III. ①企业管理—数据处理—研究 IV. ①F272.7

中国版本图书馆 CIP 数据核字(2018)第 137064 号

出版发行:中国电力出版社

地 址:北京市东城区北京站西街 19 号(邮政编码 100005)

网 址:<http://www.cepp.sgcc.com.cn>

责任编辑:刘 焜 liuchi1030@163.com

责任校对:马 宁

装帧设计:张俊霞

责任印制:杨晓东

印 刷:北京盛通印刷股份有限公司

版 次:2018 年 8 月第一版

印 次:2018 年 8 月北京第一次印刷

开 本:787 毫米×1092 毫米 16 开本

印 张:13.5

字 数:285 千字

印 数:0001—2000 册

定 价:78.00 元

版 权 专 有 侵 权 必 究

本书如有印装质量问题,我社发行部负责退换

编 委 会

主 编 吴文宣

副主编 刘永清 蔡力军 叶 勇 辛 永 黄文思

委 员 连纪文 吴 飞 王国强 罗富财 李 霆 黄 莉

上官霞 林日晖 邹玲枫 蔡宇翔 付 婷 陈 红

朱碧钦 陈 倩 刘立超 谢 毅 吕 云 蒋 鑫

杨启帆 陈扩松 章剑涛 罗义旺 李金湖 陈 坤

编 写 组

组 长 吴文宣

副组长 吴 飞 黄文思

成 员	章剑涛	罗义旺	李金湖	陈 坤	张毅琦	许梓明
	余仰淇	刘燕秋	林海玉	林永辉	刘 青	刘庆胜
	马汉斌	吴 彬	刘化龙	李雪梅	王璐鑫	林 超
	蔡 菁	翁斌新	王达琳	柳 欢	温天宝	吴胜竹
	陈 强	欧阳文佳	谢 驰	程友平	周暖青	陈 珺

随着信息技术和人类生产生活交汇融合，互联网快速普及，全球数据呈现爆发增长、海量集聚的特点，对经济发展、社会治理、国家管理、人民生活都产生了重大影响。世界各国都把推进经济数字化作为实现创新发展的重要动能，在前沿技术研发、数据开放共享、隐私安全保护、人才培养等方面做了前瞻性布局。中国政府部门印发了《促进大数据发展行动纲要》《大数据产业发展规划（2016—2020年）》等指导性文件，提出实施国家大数据战略的重大决策。

在国家电改和推动能源互联网背景下，电力企业正面对更多的技术创新和业务创新方面的挑战，应对这些挑战很大程度上依赖于对电力大数据的充分挖掘利用。电力大数据贯穿于电力产业的各个环节，如何对海量的电力大数据进行安全存储、高效计算处理和管理以支撑电力大数据挖掘变得越来越重要。因此，电网企业全业务统一数据中心作为电力数据存储和处理、业务数据融合的载体应运而生，并在大数据存储、处理、分析结果输出、可视化展示的全链路中起到了关键的承载和支撑作用，是破解企业数据共享难题的重要途径，对于推进源端业务融合，提升数据质量、增强数据共享，提高后端大数据分析应用水平，推进信息化企业建设具有重大促进作用。本书主要分为全业务统一数据中心概述、总体设计、关键技术、建设与管理及案例与实践五个部分。对电网企业全业务统一数据中心的架构设计、关键技术和建设管理实践过程进行深入剖析。

全业务统一数据中心概述部分主要介绍了电网企业全业务统一数据中心产

生背景、研究现状和发展趋势。强调了全业务统一数据中心是提升业务融合水平，深入挖掘数据价值，实现“用数据管理企业、用信息驱动业务”这一目标的关键和基础，是建设信息化企业的重要内容和必由之路。

全业务统一数据中心总体设计主要介绍了总体架构设计，并对处理域、分析域、管理域的设计思路进行了分别介绍，数据处理域是保障数据质量的关键，提升数据应用水平的基础；数据分析域是挖掘数据资源价值，提升数据应用水平的核心；数据管理域是实现数据规范、统一、安全的关键和保障。

全业务统一数据中心关键技术部分对多租户模式实现数据隔离技术、异构数据统一访问技术、异构数据采集存储等关键技术进行剖析。关键技术支撑着全业务域的正常工作。

全业务统一数据中心建设与管理部分主要介绍了建设过程中的各环节工作思路和管理内容，包括数据收集处理、差异分析、系统部署配置和集成测试、上线和建转运等环节的工作。全业务统一数据中心是公司面向全业务范围、全数据类型、全时间维度的数据集中地，提供统一的存储、管理与服务，因此，在实施过程中各环节的建设和管理工作需要与项目单位的实际情况高度匹配，并基于此制订各阶段建设方案及计划。

全业务统一数据中心案例与实践部分主要通过六个案例来介绍电网企业运维检修、客户服务、经营管理等业务领域的实践成果。剖析了电网企业充分发挥全业务统一数据中心的应用价值，以高度融合的数据为驱动，结合公司各专业部门跨专业的应用需要，提高业务创新能力和技术创新能力的典型案例，反映了电网企业全业务统一数据中心在实现业务融合、促进数据充分挖掘共享的成果。

福州大学教授

牛玉贞

国网福建省电力有限公司在运营和管理的过程中会产生大量的数据，这些数据是具有内生性和独享性的，使得电网具备开展大数据应用和商业化的基础。具体来说，主要有三个方面的优势：首先打破了过去按照业务条块各自建设、使用和维护的局面，实现了业务数据横向到边、纵向到底的贯通；其次在推进的过程中，电网各专业数据实现统一，有助于提高数据的利用率、共享水平和数据质量；最后是通过全业务统一数据中心建设，推进大数据分析深入应用。

需要注意的是，过去电网曾经积累了大量的数据，并不是所有的数据都有用。此次数据中心建设时，不能把过去的数据一股脑全装进去，而是要进一步分析，剔除一些不用的数据。这就像我们家里的书架，过一阵就要翻一翻，有价值的书保留，有些不合适的淘汰，随时保证知识的与时俱进。数据中心建设既是基础工程也是长远工程，建设过程中要对未来的发展有预判。过去电网可能采用的是直接存储数据的方式，今后要根据发展对这些数据进行深入分析，需要增加新的模块和工具。比如，智能电能表除了产生电流、电压等常规数据外，是不是可以开展用电模式分析？电网的运营数据是反映国民经济运行情况的一个重要指标，对一个地区的经济建设情况进行预测时，电网企业的售电量、用电模式、各种新能源等数据怎样转化为决策层可以利用的数据？这些都必须统筹考虑和布局。电网数据有很大一部分是可以向社会公开的，比如为经济

社会运行提供服务。同时，外部的一些数据也可以为电网企业运营提供帮助，比如天气数据，在暴雪等极端气候条件下电网需要如何保障安全运行、检修工作如何安排等。

本书从全业务统一数据中心概述、总体设计、关键技术、建设与管理、案例与实践五个方面为我们详细阐述了全业务统一数据中心在电力企业中的应用成效。也让我们对电力数据的新技术发展有了新的认识，对于在大数据公共事业及社会发展中的作用有了新的启示。

吴文宣



目录

序
前言

第1章	全业务统一数据中心概述	1
1.1	全业务统一数据中心定义	1
1.2	全业务统一数据中心研究现状	6
1.3	全业务数据中心趋势分析	13
第2章	全业务统一数据中心总体设计	15
2.1	全业务统一数据中心总体架构	15
2.2	处理域分项设计	23
2.3	分析域分项设计	26
2.4	管理域分项设计	35
第3章	全业务统一数据中心关键技术	55
3.1	多租户模式实现数据隔离技术	55
3.2	异构数据的统一访问技术	57
3.3	异构数据的采集技术	58
3.4	异构数据的统一存储技术	61
3.5	数据分析挖掘技术	64
3.6	数据可视化技术	65

第 4 章 全业务统一数据中心建设与管理 66

4.1 差异分析及方案设计 66

4.2 数据收集与处理 71

4.3 系统部署及配置 73

4.4 系统集成 78

4.5 系统测试 80

4.6 上线准备及切换 83

4.7 上线试运行支持 83

4.8 系统建设转运维 84

第 5 章 全业务统一数据中心案例与实践 85

5.1 设备资产一致性监测分析 85

5.2 配网运行智能监视 93

5.3 同期线损应用跨域实时计算 102

5.4 峰谷分时电价效益评估 106

5.5 财务经营诊断分析 107

5.6 电网规划应用 131

参考文献 202

后记 204

全业务统一数据中心概述

1.1 全业务统一数据中心定义

1.1.1 概念

数据是信息化的核心，建设全业务统一数据中心是破解企业数据共享难题的重要途径，对于推进源端业务融合，提升数据质量，增强数据共享，提高后端大数据分析应用水平，推进信息化企业建设具有重大促进作用。

全业务统一数据中心是传统数据中心的进一步发展与完善，全业务统一数据中心是提升业务融合水平，深入挖掘数据价值，实现“用数据管理企业、用信息驱动业务”这一目标的关键和基础，是建设信息化企业的重要内容和必由之路。如图 1-1 所示。



图 1-1 全业务统一数据中心的建设目标

全业务统一数据中心以“数据干净透明、模型规范统一、分析灵活智能”为具体目标，实现面向全业务范围、全数据类型、全时间维度数据提供统一的存储、管理与服务，实现业务高度融合、数据充分共享。

(1) 实现通过强化统一数据模型与企业级主数据的全面应用与管控，实现企业业务数据在语法与语义上的统一，保证数据的一致性与可用性。

(2) 实现通过改善业务集成，消除数据冗余，归并整合业务系统，实现源端数据逻辑统一、分布合理、干净透明。

(3) 实现通过全业务数据抽取、清洗和转换，构建统一数据分析服务，实现跨专业数

据的高效计算、智能分析和深度挖掘。

1.1.2 特征

全业务统一数据中心是随着企业业务的发展、大数据等新技术的应用，而逐步形成、提出的一种新数据中心建设模式，全业务统一数据中心的核⼼在于数据，数据是企业生产管理中积淀下来的资产，数据可以促进业务变革，实现架构优化，推动应用提升，因此以数据为核心，全业务统一数据中心可总结为以下四个方面特征（见图 1-2）：

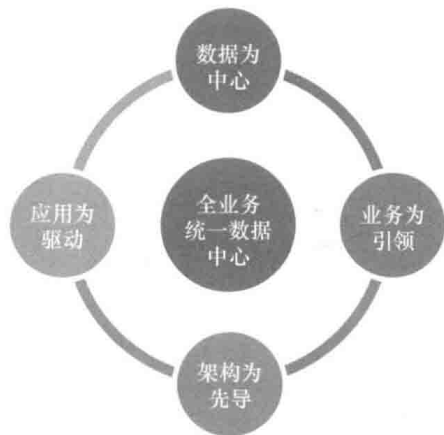


图 1-2 全业务统一数据中心四大特征

建设各个业务系统，从根本上解决数据的一致性和完整性问题，真正实现企业级的业务协同和流程贯通。

（2）业务为引领。全业务统一数据中心的建设不仅是一个技术问题，更是一个业务问题。全业务数据中心的建设方案以业务为引领，通过各个业务部门深度参与，充分分析业务需求，实现业务与信息技术的深度融合，切实解决业务实际问题，全面支持业务发展和创新。只有在业务部门深度参与下，业务与技术深度融合，全业务数据中心才能发挥出应有的价值。

（3）架构为先导。全业务统一数据中心的建设是一项庞大的系统工程，涉及专业多、影响范围大、全面可靠的架构设计尤为关键。全业务统一数据中心的建设坚持统一架构设计，一张蓝图绘到底，分步实施。通过数据规划先行，以统一数据模型和主数据管理统领整个系统的建设，在一个数据平台实现面向全业务、全类型数据的统一存储、管理与服务。

（4）应用为驱动。全业务统一数据中心作为企业统一的数据汇集平台，其建设目标是整合企业内部（生产数据、经营管理数据等）、外部数据（宏观经济数据、地理空间数据、气候气象数据等），贯通专业间、上下游数据，全面支持企业业务发展和管理决策。企业各业务部门都要主动将自己管理的业务数据和外购的第三方数据接入全业务统一数据中心，并基于此开展数据应用。

1.1.3 与数据中心的区别

按照传统数据中心的定义，广义数据中心是指企业的业务应用与数据资源进行集中、

集成、共享、分析的场地、工具、流程等的有机组合。广义数据中心的核​​心内容包括业务应用、ETL、ODS、数据仓库、数据集市、商务智能，也包括物理的运行环境（中心机房）和运行维护管理服务。而狭义数据中心仅包含应用层面的数据中心。

全业务统一数据中心是数据中心的进一步发展和完善，比传统数据中心更加强调数据与业务的融合，从源端业务系统到数据中心、再到应用决策分析，打通数据的流转、实现通过数据看业务，以数据驱动业务的发展，同时借助云计算、大数据、物联网等新兴技术，实现数据的全业务、全过程、全景式的管理与应用。如图 1-3 所示。

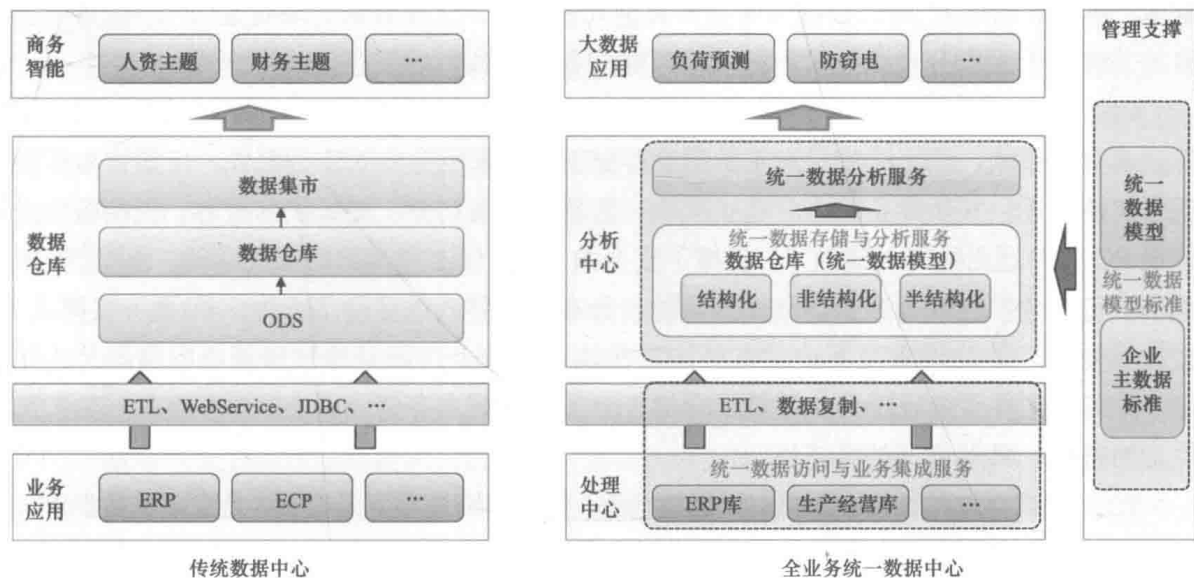


图 1-3 全业务统一数据中心与传统数据中心的区别

与传统数据中心相比较而言主要体现在以下方面：

一是全业务统一数据中心强调建立统一数据模型标准体系，以指导、规范业务应用、数据仓库及分析应用的统一建设。模型的规范统一，通过强化全业务统一数据中心统一数据模型与企业级主数据的全面应用与管控，保证数据的一致性与可共享，通过模型与主数据，实现各业务条线的信息集成与贯通，保证业务的连贯性、持续性。

二是全业务统一数据中心强调建立统一的数据访问及业务集成服务，实现对数据存储进行统一构建，对业务应用进行统一规划，实现应用与数据隔离，保证数据的干净透明，通过改善源端业务集成，消除数据冗余，归并整合业务系统，实现源端业务系统数据逻辑统一、分布合理、干净透明。

三是全业务统一数据中心强调建立统一的数据存储与分析服务，实现对全业务数据的集中存储与管理，构建统一数据分析服务，实现分析的灵活智能，通过汇总、清洗、转换全业务、全数据类型、全时间维度数据，实现跨专业数据的高效计算、智能分析和深度挖掘。

1.1.4 与新技术的应用关系

大数据、云计算等新技术日趋成熟，使得 PB 级数据统一存储与管理、跨地域分布式

计算、基础软硬件资源弹性扩展成为可能，为建设全业务统一数据中心提供了技术保障。

按照全业务统一数据中心的建设目标及业务特性，在新技术引入方面，主要涉及数据存储、数据处理、数据挖掘及分析等技术。

1. 数据存储方面

全业务统一数据中心主要面向全类型数据（结构化、半结构化、实时、非结构化）的存储、查询，以海量规模存储、快速查询读取为特征。在低成本硬件（X86）、磁盘的基础上，采用包括分布式文件系统、行式数据库（分布式关系型数据库、键值数据库、实时数据库、内存数据库）、列式数据库等业界典型功能系统，支撑数据处理高级应用。大数据存储的关键在于采用分布式技术和低成本存储设备。具体有以下存储技术方式：

（1）分布式文件系统。面向海量规模的非结构化、半结构化数据存储，传统的集中式、阵列式存储模式，已无法满足海量数据的存储需求。采用分布式文件系统，可解决海量数据存储的难题，其所具备的全分布式架构、数据块粒度切分、在线扩容减容、复制备份及普通 PC 硬件适用性等关键技术，支撑了安全的 PB 级以上规模数据在线存储，使安全、低成本、可任意扩容的大数据存储成为可能。分布式文件系统是指基于客户机/服务器模式，文件系统管理的物理资源不一定直接连接在本地节点上，而是通过计算机网络与节点相连。分布式文件系统表现为文件数据存储分散在分散的低成本存储介质上，对外提供一致的文件访问接口，具有良好的容错性。

（2）列式存储数据库。面向海量规模数据，数据的高速检索及查找，是大规模数据深入挖掘的基础。业界经典的分布式文件系统提供了数据海量存储的方法，而列式存储数据库则提供了海量数据高速增加、检索及查找的可行路径。与传统的行式存储不同，列式存储避免了维护大量的索引和物化视图，降低了时间处理及空间存储方面成本，其列存储特性，每一列单独存放，数据即是索引，只访问查询涉及的列，大大降低了系统 I/O，每一列由一个线程来处理，而且由于数据类型一致，数据特征相似，极大方便压缩，特别适合大批量数据查询。

（3）行式存储数据库。针对结构化数据，行式存储数据库其独有的小批量数据处理能力，强逻辑性、事务性及完善的 SQL 接口支持，特别适用于处理结构化数据。分布式关系型数据库完美地解决了传统关系型数据处理难题，针对特定场景的、特定数据结构的键值数据库、实时数据库、内存数据库使全业务统一数据中心能够柔性、高效的处理时序数据、键值数据、内存数据。

行式存储数据库是以行相关的存储体系架构进行空间分配，主要适合与小批量的数据处理，常用于联机事务型数据处理。目前，绝大多数数据库软件是属于传统行式数据库的范畴，满足海量规模存储、快速查询读取的行式数据库类型包括以下几种：

（1）分布式关系型数据库：通常使用较小的计算机系统，每台计算机中都有数据库的一份完整拷贝副本，并具有自己局部的数据库，位于不同地点的许多计算机通过网络互相连接，共同组成一个完整的、全局的大型数据库。

（2）键值数据库：以 Key/Value 模型为基础，主要会使用到一个哈希表，这个表中有一个特定的键和一个指针指向特定的数据。键值数据库适用于高性能半结构化数据查询

场景。

(3) 实时数据库：专门设计用来处理具有时间序列特性的数据库管理系统，针对实时高频采集数据具有很高的存储速度、查询检索效率以及数据压缩比。实时数据库支持实时、准实时数据的存储和查询。

(4) 内存数据库：将结构化数据放在内存中直接操作的数据库，相对于硬盘，内存的数据读写速度要高出几个数量级，极大地提高应用的性能。内存数据适用于高性能实时查询分析场景。

2. 数据处理方面

数据处理技术是全业务统一数据中心技术框架中的核心层，可由计算（数据查询、内存计算、流式计算、批量计算等）、数据分析算法，以及分析挖掘工具组成的数据处理功能集合。计算的核心能力是分布式计算，通过分布式计算能将一台计算机无法处理的任务分解到多个节点上。以分布式计算为核心，发展出其他计算模式构成可适应多种计算场景的计算框架。具体有以下计算处理技术方式：

(1) 内存计算技术。随着信息化水平不断提高以及信息系统不断集中，日益增大的数据量和大并发用户访问对信息系统的实时处理能力提出了更高要求。传统数据库技术在处理海量数据时，无法实现良好的水平扩展性且现有技术难以有效解决因磁盘 I/O 引起的性能瓶颈问题；随着低价的多核处理器和大量内存的涌现，内存计算将成为最终跨越大数据计算性能障碍，实现高实时高响应计算的一个最有效技术手段。内存计算技术指数据存储和计算全部位于主内存中，利用 CPU 和内存的速度和性能优势，结合并行计算技术，实现高性能计算。

(2) 流数据处理技术。企业每天的生产经营活动中都会产生海量的视频、音频、日志等流式数据，流式数据具有实时性、易失性、突发性、无序性、无限性等特征，如何从流式数据中快速地获取有价值的信息是流式数据处理的难点，系统采用流计算可以解决此问题。流计算技术是一种高实时性的计算技术，指当一定时间窗口内应有系统产生的流动数据到达后不进行存储，而是将流式数据直接导入内存进行实时计算，从流动的、无序的数据中获取有价值的信息输出。流计算具有分布式、低延迟、高性能、可扩展、高容错、高可靠、消息严格有序、定制开发等特点。

3. 数据挖掘及分析方面

在数据挖掘及分析方面，全业务统一数据中心应提供自助式的分析工具，提供易用、快速、灵活的可视化设计器和丰富多样的可视化控件，结合数据建模发布的数据主题和数据挖掘发布的业务挖掘模型，可自定义配置分析界面，方便快捷自助实现采样数据的数据透视分析、报表制作和图形展示，为业务使用数据展示数据价值，提供应用思路，为业务应用的敏捷开发提供支持。具体有以下计算处理技术方式：

(1) 数据挖掘技术。数据挖掘是一种决策支持过程，它主要基于人工智能、机器学习、统计学技术，通过对原始数据自动化的分析处理，做出归纳性的推理，得到数据对象间的关系模式，这些关系模式反映了数据的内在特性，是对数据包含信息的更高层次的抽象。与侧重于根据已知训练数据做出预测的机器学习相比，数据挖掘更关注于发现未知的信

息。数据挖掘任务可以分两类：描述性和预测性。描述性挖掘任务刻画数据的一般特性，主要基于非监督学习算法；预测性挖掘任务在当前数据上进行推断，以进行预测，一般基于监督学习算法。

(2) 自助式分析技术。自助式分析技术可提供易用、快速、灵活的可视化设计器和丰富多样的可视化控件，自助式分析服务则结合数据建模发布的数据主题和数据挖掘发布的业务挖掘模型，使全业务统一数据中心用户可自定义配置分析界面，具体应包括分析场景发布、可视化设计、业务建模等技术能力。

1.2 全业务统一数据中心研究现状

1.2.1 电力企业研究现状

1. 研究背景与目的

数据是信息化的核心，“数据即资产”已经成为当前业界一个普遍共识。国家电网公司一直以来高度重视信息化建设工作，经过 SG186 和 SG-ERP 工程的建设和应用，国网公司已经建成总部、省（市）公司两级数据中心，积累数据总量超过 5PB，月增长量约 46TB。数据范围覆盖各行各业、千家万户，是一座价值巨大的“金矿”。如何对企业已有的海量数据进行统一的管理与应用，充分发挥数据价值，是当前公司发展迫切需要解决的问题。

回顾近十年的信息化发展历程，各阶段的特点不尽相同，但一脉相承。“十一五”阶段，也就是 SG186 建设时期（现在统一称为 SG-ERP1.0），主要是“填补空白式”的信息化建设，主要解决“从无到有”的问题，当时的数据模型是 1.0 版本，主要服务于各个专业的信息系统建设，数据中心的建设和应用处于初级阶段，重点是满足业务应用功能，共享集成需要还较少；“十二五”阶段，也就是 SG-ERP 建设时期（现成统称为 SG-ERP2.0），主要是“集中集成式”的信息化建设，主要解决“从有到优”的问题，数据模型从 1.0 版本发展到“十二五”末，逐步形成了 2.0 版本。同时，开始推行主数据管理，逐步推动统一编码的应用，主要服务于各专业间的集成共享信息化建设需求，数据中心随着发展到公共数据资源池，这个阶段的 2.0 数据模型是共享模型。

进入到“十三五”，也就是目前正在开展的 SG-ERP3.0 建设，各专业、各单位的信息化建设和应用不断深入，跨专业协同需求成为普遍现象，“用数据说话、用数据决策、用数据管理、用数据创新”的需求日益广泛而迫切。这些需求均离不开信息化建设的进一步提升和发展，主要体现在对数据的采集、存储、计算、分析和管理提出了更高要求。这就迫切要求尽快实现三个方面的转变，一是数据模型从共享式的 2.0 版本向全业务的 3.0 版本转变，从仅仅应用于分析决策类领域向同时应用于分析决策和业务处理两类领域转变；二是主数据管理和应用从局部编码统一向企业级应用和实时管控转变；三是以上述三个转变为基础，加快推动数据中心由“共享式、分析型”数据中心向全业务数据中心转变。

通过全业务数据中心建设，可以大幅提升公司各专业、各单位数据综合利用水平。例