

大数据环境下企业数据管理存在的问题及优化策略

路欣¹ 张嘉希²

1. 国网河北省电力有限公司信息通信分公司, 河北 石家庄 050000

2. 国网河北省电力有限公司宁晋县供电分公司, 河北 邢台 054000

摘要: 数据要素价值的释放, 在很大程度上取决于企业的数据管理水平。目前大部分企业虽然拥有了大量的数据但是又陷入了数据孤岛、数据质量失控和安全合规风险的问题当中。“沉睡的数据”数量众多。因此, 本文对大数据环境下企业数据管理中存在的问题进行了深入的研究(比如: 数据资源分散难以有效整合; 缺乏有效的数据质量管控体系; 面临突出的安全与合规风险; 存在数据管理的组织与人才短板等)并以此为基础提出了从构建企业级数据中台、建立全生命周期的质量管控体系、完善的安全治理体系、健全组织人才保障以及完善智能化治理等方面来提升企业数据管理水平的相关建议, 希望可以对企业由“管好数据”向“用好数据”的转变起到一定的借鉴作用。

关键词: 大数据; 数据管理; 数据治理; 数据孤岛; 数据质量

0 引言

数据已经成为新的五大生产要素之一。据 IDC 统计, 在 2025 年时中国产生的数据将达到 51.78ZB, 但是只有不到六分之一的数据被真正保留下来, 大量的数据由于缺乏有效的治理而成为“沉睡资产”, 根据信通院对企业的调研发现有 78% 的企业已经开展了数据治理工作, 但却仅有不到 3 成实现了真正的数据资产化的运营、“高投入、低产出”的现象十分突出; 目前企业在数据管理方面存在诸多的问题: 如技术标准混乱、组织职责不清、安全合规的压力增大等等。所以需要找出其中存在的突出问题, 并给出可以落地的优化建议是很有必要的。

1 大数据环境下企业数据管理的突出问题

1.1 数据孤岛与整合效率低下

数据孤岛是企业在进行数据管理过程中最为具有普遍性的难题之一, 每个业务部门都有自己单独的数据存储体系, 如: 销售部门有自己的客户数据库, 生产部门有自己的生产数据系统等, 不同部门之间的数据标准和格式并不统一, 并且彼此间存在信息壁垒, 难以打通。据相关研究显示, 在一个企业当中, 元数据分散于各个不同的异构系统之中, 需要花费 3 到 5 天的时间才能完成对企业所有元数据的盘点工作, 其准确率不到 70%; 数据血缘断裂率为 35%-40%, 人工对变更的影响进行追溯要耗时 2-3 天。这就造成大量的数据不能被集中起来使用, 当其他部门想要获取某项数据的时候, 往往需要消耗大量的人工来整合这些数据, 不仅增加了企业的管理成本, 也耽误了使用的最佳时机。同时, 越来越多的外部数据源使得企业和外部数据缺少有效的对接方式, 更不利于整合。由于数据资源过于分

散, 企业很难拥有全面而准确的数据视图, 直接影响着数据驱动决策的实现。

1.2 数据质量管控体系缺失

数据的质量问题是价值释放的前提条件之一, 但是大多数企业的质量管控还停留在被动应对的状态上, 在很多大数据项目中因为数据输出质量不高而导致没有实现预期的目标。究其原因主要有以下几个方面的问题: 从标准层面上看, 并不存在一套统一的数据编码规则和命名规范, 同一个字段在不同的系统中有几十个定义, 使得数据不能有效地进行关联; 从流程层面上看, 对数据质量的检查一般都发生在数据生产之后、“以用促改”, 问题出现得比较迟、修复的成本比较高, 有些问题根本就没办法彻底地修好; 从工具层面上来看, 靠人工去排查非常低效且不现实, 面对海量数据需要实时监控的情况更是如此。并且数据质量问题会产生一定的传导作用, 一个上游系统的质量问题会随着多个级别的流转加工过程逐渐被逐级扩大化, 最终影响到分析结果的可靠性。许多企业在数据仓库和 BI 系统的投资巨大, 但却由于底层数据存在的质量问题造成分析结果不可信, 从而降低了数据的应用价值。

1.3 数据安全与合规风险凸显

一是大数据环境下的数据安全及隐私保护问题。一些企业对于数据的安全性不够重视, 在存储、传输、使用等方面缺乏足够的安全保障, 容易受到黑客攻击和发生数据泄露的问题, 一旦企业的核心商业数据或者客户的个人隐私信息被窃取, 会直接造成经济损失以及影响其信誉度。从近年来的执法情况来看, 未建立数据安全管理制度、未采取相应的技术措施保障数据安全是比较常见的现象。更为深层次的是技术和管理上的困难和技术

难题。一方面,数据在采集传输过程中可能产生泄露风险,也可能出现违规操作、越权访问等情况;另一方面,目前的数据分类分级标准还不完善,授权边界不清,权限控制粗放等等都成为存在的突出问题。同时,《数据安全法》《个人信息保护法》等一系列法律法规深入推进之后,企业在数据收集、存储、使用和共享各个环节都会面临着更加严格的要求,合规的压力越来越大。

1.4 数据管理组织与人才短板

有效的数据管理需要有良好的组织保证以及专业的人才支持。但很多企业的数治理还处于“受管理级”,存在着战略执行不闭合、责任落实不到位、数据规划不足的问题;一方面,在组织上,数据管理的工作职责被分布在各个不同的部门,没有形成一个统一的治理委员会,也没有清晰的责任划分,因此对于决策很难达成共识,执行也难以真正落地。另一方面,人才方面,管理人员数据素养和应用能力不足已经成为制约大数据有效应用的重要瓶颈因素。许多管理者是在传统的管理环境中成长起来,并不具备系统的大数据知识储备,无法从海量数据当中挖掘出有价值的有用信息,在实际工作中依旧依靠经验和感觉进行决策,对企业推行数据驱动的管理模式持怀疑态度;并且,目前大多数的企业都缺少面向管理人员开展数据素养方面的培训体系,既懂业务又懂数据的复合型人才极度稀缺,使得数据治理工作难以持续深入开展。

1.5 治理与业务“两张皮”

很多企业在开展数据治理过程中会遇到一个很常见的问题——数据治理“建而不用”。不少企业的数治理项目建设之初就非常热闹地成立了治理组织、梳理了数据标准、建设了元数据平台、制定了质量规则、盘点了数据资产,在项目验收的时候看起来一切都很完美;但是经过一段时间的运行之后发现,标准没人查,元数据没有人管,还是得靠人工去排查质量问题,资产目录建成以后也没有人去看,数据治理平台变成了“只有治理部门在用”的后台系统。究其原因是治理跟业务脱节的问题。因为数据治理的关键是业务和技术协同,工具只能作为辅助手段,并且如果没有清晰的业务定义以及流程规范,再好的工具也无济于事。如果不能把治理的工作同业务场景联系起来,则业务部门不会感受到治理带来的价值,参与的积极性就会下降,治理工作就没有动力继续下去。

2 大数据环境下企业数据管理的优化策略

2.1 构建企业级数据中台,打破数据壁垒

要解决数据孤岛的问题,首先要有一个统一的数据整合和服务平台,在这个平台上可以做到打破不同部门之间数据壁垒,

统一数据的标准及格式并且企业内部外部的各种类型的数据汇集起来、存储以及治理的工作。在技术方面,本文可以采用一种分层次的方式搭建数据中台,其中数据采集层使用 Flume 和 Kafka 来完成对各种来源的不同类型的海量数据的快速高效的接入,治理层使用 Spark 和自定义校验引擎等工具来进行数据标准化处理并修正质量问题,存储层本文可以通过 Hive 和 HBase 两种方式相结合的方式来满足离线分析和高并发查询的需求,而服务层则是提供 RESTful 和 GraphQL API 给上一层的应用调用数据的能力。

在建设路径上不要追求“大而全”的一次性投入,从小切口着手不断迭代;也可以先围绕客户、产品、财务等主题域做数据整合的试点工作,验证效果后再逐渐扩大范围。但是也要注意,数据中台的建设不能仅仅停留在技术层面上,同时要做数据标准的制定以及治理流程的设计等工作,不然就会导致数据从“分散”变为了“集中式的混乱”,反而治理的效果会更差。

2.2 建立全生命周期数据质量管控体系

要实现由“事后补救”的方式转变为“源头治理”,在标准上,企业应该根据国家标准以及行业最佳实践的基础之上,按照自身的业务特点来进行统一的数据编码规则、命名规范和属性的定义,并且建立一套包含完整性、准确性、一致性、时效性、唯一性等质量评估指标体系来利用自动化工具对数据质量进行监测。

在流程上应该把质量检查嵌入到数据采集、加工、流转过程中的关键环节,在事前进行事中校验并且在事后进行监测。通过在中国银联实际应用发现:建立数据质量全流程嵌入式管控体系是能有效的让质量管理从被动应付转变成积极预防;在技术方面,深度学习方法相比于传统的规则方法,在异常检测、数据去重、缺失值填补以及实时数据清洗方面的性能要好于传统的方法,企业可以引入基于自编码器、Transformer 等模型智能化的质量管控工具来提高管理和效率。

2.3 完善数据安全治理体系

数据安全治理应该覆盖到整个数据生命周期中:从技术角度采用数据加密、访问控制、入侵检测等手段加强数据存储、传输和使用的安全保障并且可以利用可信执行环境实现“数据可用不可见”,用智能合约实现权限的自动管控;从制度层面对数据进行分类分级管理和明确数据安全的责任划分,对于数据收集、使用及销毁流程做出规定;从人员层面展开针对员工的数据安全培训工作以提高其安全意识,避免由于个人原因造成的数据安全问题发生。

值得指出的是，在数据安全治理方面，“成本负担”的标签正在被“价值源泉”所取代。随着合规基础设施的落地和标准的国际化，数据合规的能力正逐渐变成企业进入数据要素市场进行交易的基本要求，那些率先建立起完善的合规体系的企业将会在竞争中占据先机。

2.4 健全组织保障与人才培养机制

数据管理的不断运转离不开组织、制度、流程的支持，在组织上，企业应该成立一个统一的数据治理委员会，并且由高层领导牵头，明确数据 Owner 和数据 Steward 的职责划分：数据 Owner 是某个领域内的数据资产负责人，保证该领域内数据定义清楚并且被正确地使用；而数据 Steward 则是负责日常的数据管理工作（比如标准执行、质量和问题）等等。

在制度方面：在企业内部应该首先制定出数据治理的组织与职责管理办法、数据标准管理办法、数据质量管理办法和数据安全分类分级管理办法等相关的管理制度并对每一项治理活动责任主体、操作流程、考核标准进行规定；在人才方面：要针对不同的层级的人才开展有针对性的人才培养工作，例如对于管理层开展数据素养方面的培训，让他们了解数据的价值以及如何用数据来辅助管理者的决策；对于业务人员开展数据技能方面的培训，使他们具有数据应用能力以及自助分析的能力；对于技术人员开展数据治理的专业培训，培养既懂技术又懂业务的技术骨干等等。

2.5 推进智能化数据治理转型

人工智能正在改变着数据管理的底层逻辑：在 Gartner 看来，到 2028 年将有 33% 的企业会利用 Agentic AI 来进行数据分析和决策（而是在 2024 年的这个比例还不到 1%）；在数据治理方面，

用自然语言处理技术从业务文档中自动生成数据的标准定义、使用机器学习算法自动发现问题并推荐解决方案、利用知识图谱技术构建数据血缘关系图谱以快速分析变更的影响等等都是数据智能化的体现。

在治理模式上：由依赖人工的规则驱动转向智能驱动的方向发展；随着垂直领域应用落地，利用大模型的能力可以解决通用大模型的数据治理中出现的“知识肤浅”、“幻觉频发”的问题，通过对于深度行业的知识进行嵌入，从而实现从数据标准制定、质量管控全过程的智能赋能。企业要密切注意这个趋势的发展，并且要在时机合适的情况下去引入智能化治理工具，降低成本并提升治理效率。

3 结语

大数据环境下企业数据管理优化是一项复杂系统的工程，涉及到技术和管理、组织制度和人才能力建设等全方位的变革创新：从问题来看，数据孤岛、质量问题、安全风险和治理与业务脱节等一系列难题互相纠缠在一起是制约企业发挥数据价值的重要瓶颈；而从优化路径来讲，需要以建设统一的数据中台作为技术基础，以建立全生命周期的质量管控为核心抓手，以完善的安全部署体系作为合规底线，以健全的组织人才保障作为持续动力源泉，以推动智能化转型作为未来趋势指引——但这 5 个方面不是孤立存在的，它们是彼此之间相互支撑的有机整体。只有把数据管理上升到企业的战略能力层面，才真正能够实现从“管好数据”跃升到“用好数据”，让数据从“沉睡资产”变成驱动企业高质量发展之“核心引擎”。

参考文献：

- [1] 周书人. 大数据驱动下企业人力资源绩效管理变革的逻辑与优化路径[J/OL]. 中国企业管理知识仓库, 1-8[2026-09-20]. <https://doi.org/10.27006/j.cnki.qygy.20260915.002>.
- [2] 江聃, 郭博昊. 让统计数据更真实反映地方经济活动[N]. 证券时报, 2026-09-17(A02).
- [3] 刘溶. 企业数据资产价值实现机制及运营管理优化策略研究[J]. 经济研究导刊, 2026, (17): 91-94.
- [4] 林永悦. 大数据时代集团企业数字化转型策略探究[J]. 现代营销, 2026, (25): 118-120.
- [5] 朱怡沁. 数字化时代国有企业资金管理策略研究[J]. 现代营销, 2026, (25): 34-36.

作者简介：路欣（1986—），女，汉，河北衡水，国网河北省电力有限公司信息通信分公司，硕士，高级工程师，主要研究方向为数据管理、数据应用及人工智能。张嘉希（1988—），男，汉，河北石家庄，国网河北省电力有限公司宁晋县供电分公司，本科，工程师，主要研究方向电力系统调度自动化相关。