

煤矿数据治理特点及发展路径分析展望

韩培强 冯智愚 王海 徐永刚 李仁飞

煤矿行业数据管理的现状

2020年3月，国务院发布的《关于构建更加完善的要素市场化配置体制机制的意见》中首次提出五大要素领域，即“数据”与“土地、劳动力、资本、技术”是同等地位的社会发展要素，由此可见数据在现代国民经济中的重要地位和作用。对于煤炭企业来说，要完成企业数字化转型，关键概念是数据到信息、再到数据资产的转变过程，核心步骤是实现传统的“数据资源管理”到“数据资产利用”的转化。

在2020年以前，煤炭行业整体都缺乏对于智能化矿山整体框架的认知，尤其是数据框架的引领和顶层设计，仅是强调安全监控、人员定位等一些单系统的建设内容，这些系统仅从煤矿企业—煤炭集团—行业监管部门等纵向作出数据联动的规定，导致大部分煤炭企业在数据上呈现“纵向部分能通、横向基本不通”的“孤岛”状态（图1），严重制约了企业从数据角度发力，实现精益经营的发展诉求。

煤矿行业数据治理现状分析

近年来，随着智能化被国家和行业逐步重视，一些大型煤炭企业建立了综合管控系统IOC应用（图2），实现了一定程度的数据汇聚，在数据治理走入煤矿行业之前，大体上可分为以下情况：

（1）无IOC应用

最原始的状态，各系统孤立运行，系统之间没有联接，也没有统一的数据管理通道。

（2）通过综合自动化系统实现部分IOC应用

仅通过工业自动化软件，实现一部分自动化监测、控制应用及数据的统一，但未与其他系统融合。

（3）通过智能化矿山综合管控系统+独立数据采集程序实现IOC应用

这种情况下，建有覆盖煤矿大部分应用的“一张图”管控程序，但是在数据采集部分使用的是多个独立的数据采集程序来和煤矿已建各类系统对接，并且将所有数据统一写入一个关系型数据库当中。上述分类中的（1）（2）已经无法满足煤矿海量数据

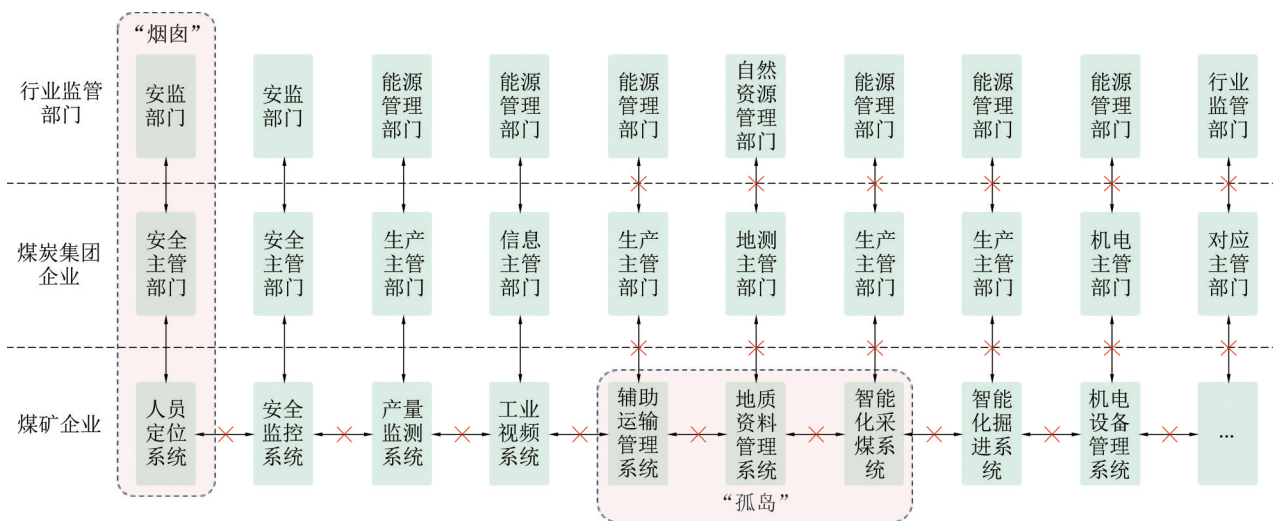


图1 煤炭企业数据孤岛状态

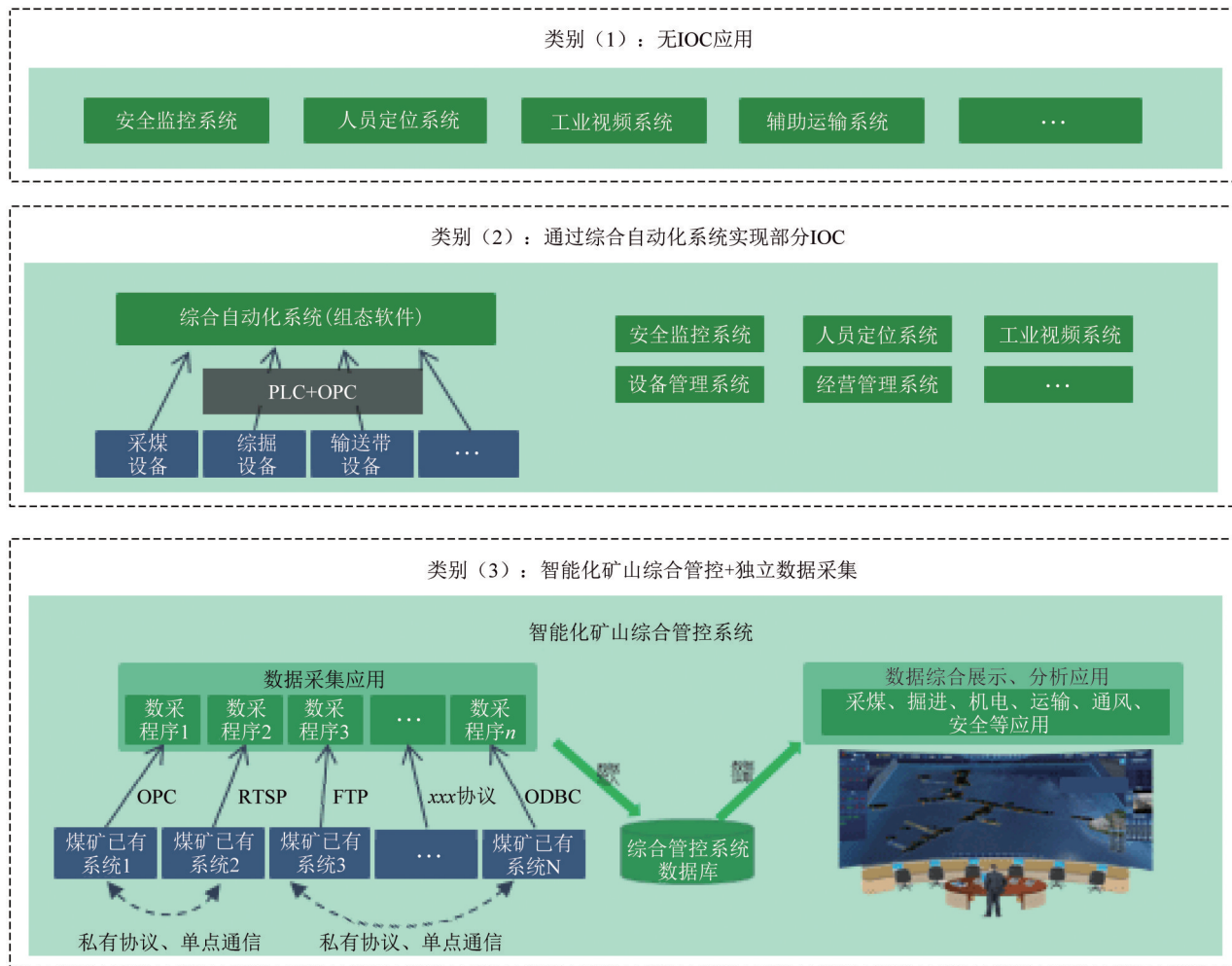


图2 综合管控系统IOC应用

和系统管控的需要，（3）中也存在诸多问题和瓶颈，主要体现在：

（1）不能适应变化

接入已建各子系统需要按不同的协议单独编写程序，建设成本高，环境发生变化时，需要修改程序，适应性差。

（2）性能瓶颈

对接煤矿数十个系统，拥有统一的采集和数据出口，用普通数据库作为承载，未考虑分布式、并发等设计需求，容易出现丢失、卡顿等性能问题。

（3）浅层次的数据集中、缺乏深层管控

没有数据标准、数据质量、数据目录、提取等工具，缺乏由数据资源管理到数据资产利用的质变条件。

（4）缺乏统一服务出口

系统之间单点对接，耦合性强，一方出问题则

无法继续，不易排查故障。

上述问题直接导致出现以下情况：

（1）系统融合难

“烟囱”与“孤岛”林立，跨业务联动困难，接口格式、通信协议各式各样，数据融合严重依赖建设商。

（2）运行效率低

传统数据库结构无法满足智能化矿山数据日益增长的需求，矿山IOC应用作为复杂巨系统，数据存储及写入成为性能瓶颈。

（3）摸清家底难，有集成无标准

没有数据资产概念和管理工具，数据如散沙，状况不明，质量不清。

（4）数据价值低，信息无专题

应用基本以展示查询为主，信息无关联、无提炼，“只见树木不见森林”，难以用于业务决策。

在智能化飞速发展的时代，海量信息和数据力量巨大且难以治理，系统建设越多，给企业造成问题就越明显，甚至使部分用户质疑数字化、智能化的必要性。

煤矿数据构成和治理需求分析

结合行业特点，矿山数据可以从业务、类别、形式上着手进行分类，如图3所示。

按业务维度，可以分为生产数据、安全数据、专题应用数据和经营管理数据；按类别来分，根据数据的持久性，可分为主数据（基础数据）和业务数据；根据时效性，可分为实时数据和历史数据；根据结构特性，又可分为结构化和非结构化数据；按格式来分，可分为直接来自工业系统的各类工业接口、关系型数据库、非关系型数据库、OPC服

务、程序API、文件服务、MQTT、HTTP、SIP、RTSP、日志数据、GIS和BIM等。只有明确数据的内容和形式，才能为下一步的数据治理理清思路，明确目标。

从数据管理建设需求来说，数据治理是煤矿智能化进程的必由之路，当前煤矿数据建设的需求主要体现在以下5个方面：

(1) 建立数据规范

需要建立统一标准规范的数据体系，规范主数据、数据索引格式、元数据格式、数据表结构、布局方式、存放格式、精度要求、时效设置和编码方案等，其中元数据和数据索引主要包括各类数据概述、用途、存放路由、数据库、访问引擎和索引结构等，体现数据的层次结构。

(2) 统一数据采集

智能化煤矿应建设大数据服务中心，统一数据

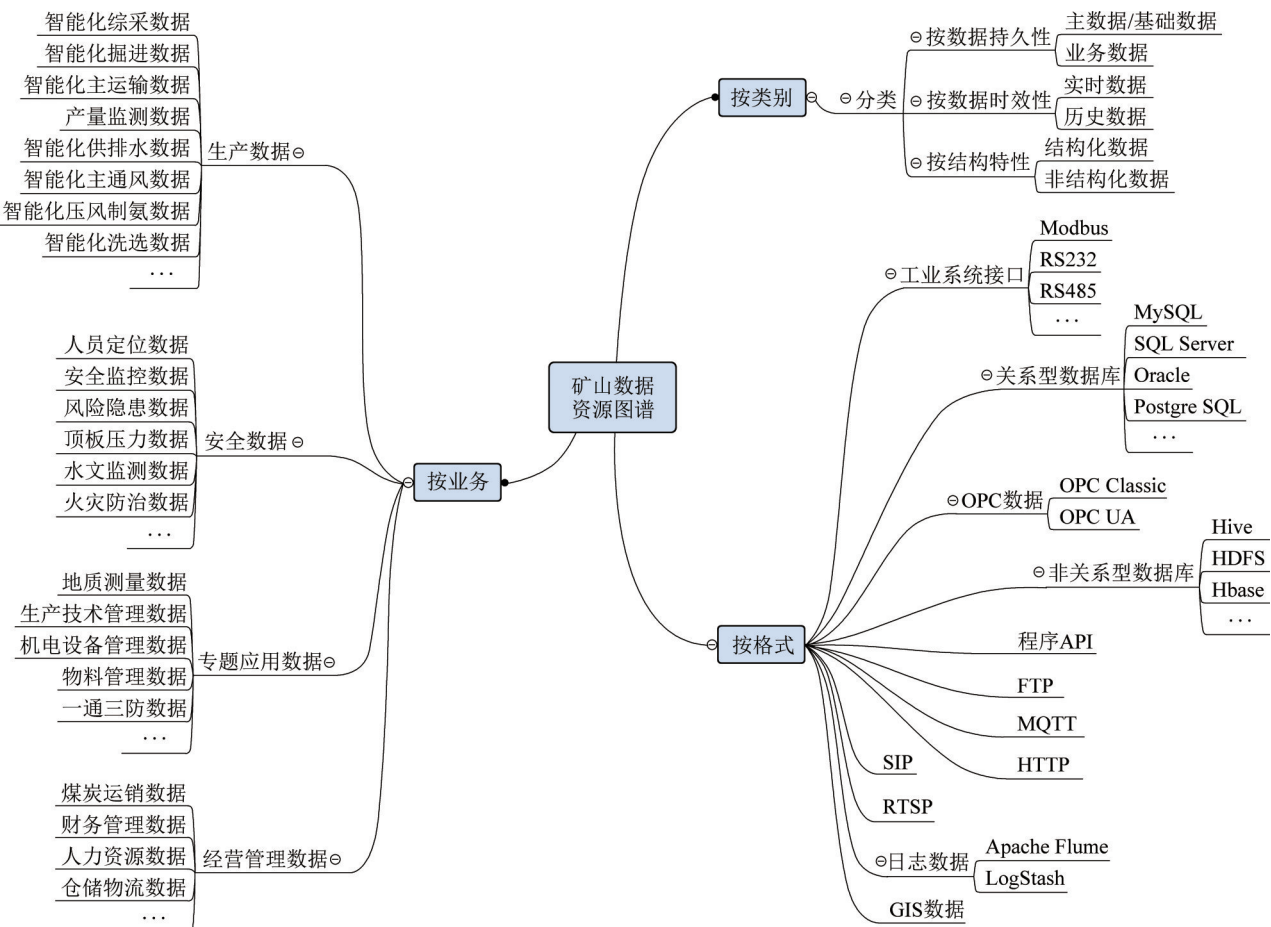


图3 矿山数据资源图谱

采集、传输、存储和访问接口标准，构建煤矿数据治理体系。能够支持多种数据服务、通信协议和接口，从各类仪表、模块等多种软件、硬件中获取数据，并能够通过开放接口向各种应用提供数据；能够从各种服务系统、应用系统和控制端获取命令，并能够自动转发和执行命令，保证命令的可靠性与时效性；能够不影响各业务系统正常运行的前提下，将产量监控、人员定位、应急广播、生产系统、经营系统、安全管控系统、设备管理系统、工业视频、通风、排水、运输等各大系统集成到综合管理平台。

(3) 建立数据管理机制

需要建立完善的数据质量管理组织架构，明确数据权属、管理者、使用者等，运用新一代信息技术建设业务中台和数据中台，为上层业务应用提供统一的数据汇聚与技术支撑。

(4) 形成质量管理手段

制定规范的数据质量改善流程，形成面向多样化煤矿数据应用场景的数据质量管理闭环。

(5) 统一共享与服务

建立统一的数据服务接口、信息采集标准、数据格式、通信协议，实现数据的统一集中管理，建立矿井多源异构信息数据共享平台。能够汇聚企业

内部各种数据资产，包括应用、服务和相关集成，支持数据共享，具备提供数字化资产运营的分析能力。

煤矿行业数据治理技术架构

煤炭企业数据治理的本质是，将原始状态分散的数据经过逐级加工，最终形成企业数据资产的转变过程，从而实现企业的“数据资源管理”到“数据资产利用”的跃变，从下往上，信息的核心加工链共分为“数通、数聚、数治、数融、数信、数享”6个环节，具体技术架构如图4所示。

“数通”

“数通”环节实现源数据从技术以及组织管理上的互联互通，是实现后续步骤的基础，对于煤炭企业来说，需要从以下角度入手解决“数通”问题：

(1) 打通数据组织

需要从数据源头的组织层面，包括数据的产生者和拥有者（主要是负责各系统的业务部门）获取数据，并建立数据的自动更新机制，以获得稳定可靠的数据来源，当然，该步骤也与组织和管理制度息息相关。

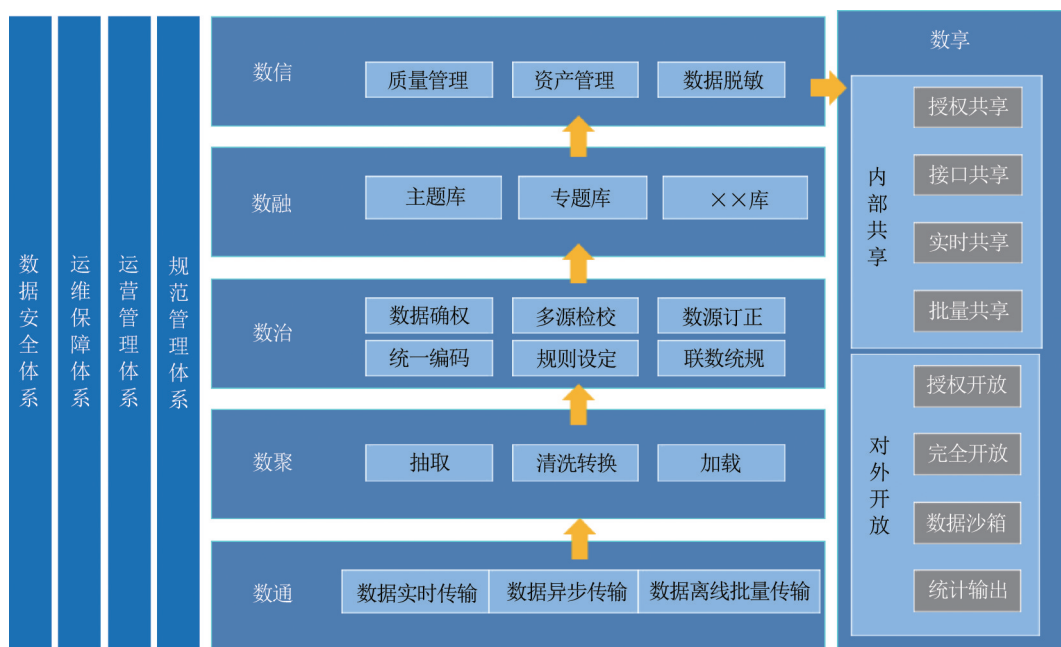


图4 信息的核心加工链具体技术架构

(2) 打通网络

针对煤炭企业的工业控制区、办公网络、企业经营管理区等不同网络域,在相互隔离的前提下,通过建立前置采集节点等手段,打通网络域,实现数据的自动更新渠道。

(3) 打通接口

对于自动化获取的数据,需要实现数据中台采集侧与系统接口方的采集途径及端口,实现对接的可行性。

(4) 打通填报渠道

对于无法通过系统和网络稳定获取的数据,采用定期填报机制,提供填报页面及导入程序。

“数聚”

“数聚”环节同样也称作数据采集,用于实现将多种不同来源、形式的数据自动化汇聚至数据库,以及数据的归一,其本质是通过ETL作业对采集数据源进行处理并做轻度汇总,分担大量的基础数据处理任务和数据存储压力,并将处理结果传递给数据仓库。

数据采集架构设计主要包括抽取(E, Extract)、清洗转换(T, Transform)、加载(L, Load)3个部分。数据抽取是指从数据源获取所需数据的过程,会过滤掉目标数据集中不需要的源数据字段或数据记录。数据清洗和转换主要包括格式转换、字段合并与拆分、数据翻译、数据匹配、数据聚合及其他复杂计算。数据加载就是将从业务系统中抽取、转换后的数据加载到数据仓库系统中。

为了应对多源的数据采集,数据采集工具应满足如下功能:①支持丰富的数据源类型,支持丰富的数据转换能力;②支持第三方kafka客户端直接对接消息集成的消息生产和消费;③支持第三方系统开放API、数据API、函数API、数据文件的接入。

“数治”

“数治”环节主要针对不同来源的“脏”数据的重复、缺失、错误等问题,按照标准规则进行治理,同类型数据进行标准化,从内容上来说分为4个模块:

(1) 元数据管理

元数据是用来描述数据的数据,它可以帮助数据管理员非常方便地找到其所关心的数据,元数据可按用途不同分为技术元数据和业务元数据。

(2) 数据标准管理

对煤炭企业生产、安全、经营与管理活动中所涉及与产生的关键基础数据和业务指标建立统一定义、解释口径与标准,完整的数据标准体系包含基础数据标准和业务指标标准2个方面的内容。数据标准体系建立必须遵循唯一性、稳定性、前瞻性、准确性、可执行性、低风险性、可实施性等设计指导原则,参考国家相关行业标准、地方标准并结合企业自身需求构建。

(3) 数据模型管理

数据模型管理的目标是提供数据平台数据模型的整体设计思路,以及各个实体数据流动规则,为项目实施数据平台数据模型提供设计指导,典型的代表是以事实表和维度表构成的星形模型。

(4) 数据血缘管理

血缘管理是指以某一实体作为起点,往回追溯其数据处理过程,直到数据源接口。血缘分析的输出结果包括回溯过程中找到的所有元数据对象,以及这些元数据对象之间的关系。通过提供可视化血缘分析界面,辅助使用者清晰地查看元数据的上下游关系,帮助业务快速定位数据问题。

“数融”

“数融”环节又称数据建模,是针对标准数据进行指标关联和进一步梳理、融合,将数据重组,形成面向不同价值的数据源,“数融”环节将生成以下不同层次的数据库:

(1) 基础库

基础库是从煤矿各业务系统、设备侧等获取的数据,参考数据治理阶段的数据规范要求,对数据按照属性划分、整合、标准化数据清洗等操作,最终建设成统一规范的基础库数据,基础库也称贴源层,紧密面向原始数据。

(2) 主题库

主题库是在基础库的数据基础上，按照业务场景，进行数据整合、轻度汇总、算法标签，面向应用提供智能数据服务。对于煤炭企业，主题库应包含组织、人员、设备、巷道、环境、运输等各类信息的分类要求。

(3) 专题库

专题库是在基础库和主题库的基础上，对数据进行统计、分析、计算、分层形成面向业务的数据仓库，并建成配套数据管理体系。专题库面向安全、生产、经营管理等不同领域，各单位可结合实际业务需求建设特色专题库。

在构建分层数据模型的基础上进行数据开发，涵盖数据管道、多维分析、实时计算、业务流程、数据建模以及时空分析、全文检索等能力集，打造出全域数据开发和全链路的数据监控，让用户能够看到整个开发链条上每个节点的开发状态和统计监控，最终通过数据地图、数据质量、数据安全等，方便企业掌控数据资产，为发展决策提供依据。

“数信”

“数信”环节主要实现对数据沉淀成资产化和安全化管理，使得数据可信可靠，本环节包含3个层面：

(1) 质量管理

数据质量管理体系，旨在为煤炭企业建立统一的数据质量度量、检核体系，全方位管理数据平台

的数据质量，帮助企业提升数据质量。数据质量的内容包括完整性、规范性、一致性、准确性、唯一性、关联性等，或者是用户自定义的数据质量指标。

(2) 资产管理

数据资产管理是为用户建立统一的数据资源目录，将数据从资源视角提供管理。

(3) 数据脱敏

数据脱敏是在保留数据原始特征的前提下，对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的保护。通过定义脱敏策略，针对不同的使用单位和不同类型的数据进行脱敏。

“数享”

“数享”环节是实现对沉淀的数据资产进行按需共享，提高数据使用价值，对于煤炭企业来说，数据资产是企业的核心资源，需要建立数据资源管理平台，实现对于各类数据的权属者、使用者的有效管理，现有各行业的数据治理案例大多以数据服务的形式向外提供数据共享接口，主要包含以下功能：①数据服务管理，包括数据服务的注册、发布、注销、订阅等；②数据服务管理，支持数据服务的增删改查，服务信息的查询、统计等；③数据鉴权，支持多种鉴权方式，完成对第三方身份的鉴权、服务本身的鉴权等功能；④数据加密，支持多种加密手段、多种加密算法，并且支持可配置的方式进行数据加密；⑤数据服务监控及日志，数据服务状态的监控及日志，记录数据重要访问信息。

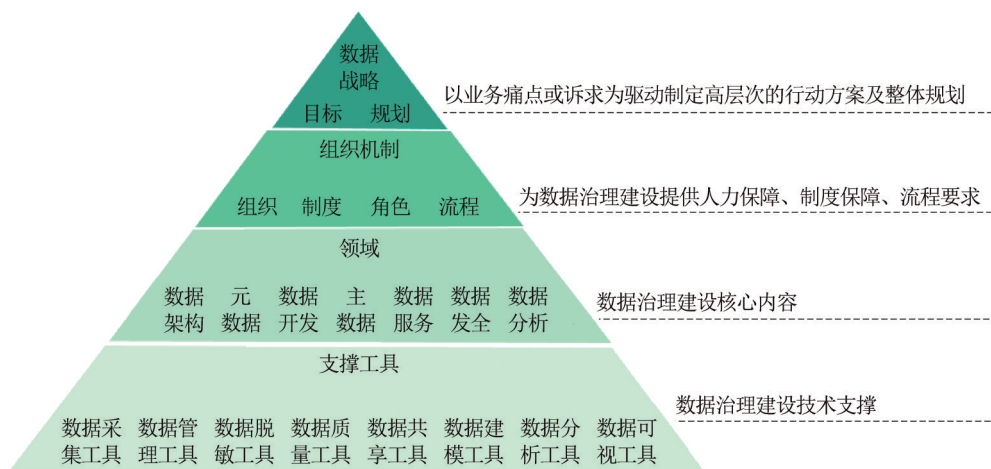


图5 煤矿数据治理金字塔架构

表1 数据治理角色分工

角色	角色定位	具体职责
煤炭企业高管	领导者	组织支撑:从组织及政策层面提出要求 推动:带动生产、机电、通风、地测等部门配合 要求:建议矿长牵头,至少副矿长级
信息化建设部门	组织实施者	标准建设:参考国家相关数据标准,建立企业数据标准 建设推进:建立数据资产管理制度、组织及工作规范,牵头推进各相关单位执行 执行:约束子系统建设商按矿方规则执行对接及上传工作
数据治理软件建设商	支撑者	服务:提供数据中枢建设咨询服务 建设:建设数据中枢核心软硬件系统 指导:全过程为煤矿用户提供建设过程的技术指导、数据梳理人员支撑 运维:完成数据资产初始化,执行运维支撑
各业务单位	执行者	配合:配合信息化建设部门进行数据梳理,协调数据提供方提供相应支撑 提出需求:以智能化为标准,作为数据使用方提出对数据资产的利用要求 使用:转换思想,以应用驱动建设
煤矿各子系统建设商	配合者	执行:按照矿方制度执行规则 原则:新建信息系统,必须要求在完工前实现与数据中枢的对接,已建系统也要积极完成对接工作

技术与管理共治 工具与组织并进

上述分析更多的是从技术角度分解,煤矿数据治理工程的实施是一个系统工程,不能只通过技术角度来实现,从组织、管理与实现角度来说,煤矿的数据治理可体现为一个金字塔架构,如图5所示。

自上而下,是数据治理的推进和分解过程,自下而上,是逐级支撑的过程。第1层是数据战略层,煤炭企业首先需要清晰地意识到自身对于数据的诉求,以及如何通过数据解决自身的业务痛点问题,制定明确的战略规划,尤其对于高层领导来说,要有清晰的意识和坚定的意志;明确目标和规划后,第2层是组织和机制层,通过确定数据治理的组织机构,明确相应的角色、制度和流程来保障整体工作的顺利实施;第3层是领域层,即实施数据治理的核心内容,对数据进行全面分析,明确各类数据架构、元数据、主数据等,进行数据安全设计,规划各类数据服务,对数据进行全面开发;第4层是支撑工具层,即数据治理所需要的各类技术及软件工具等等。具体的数据治理角色分工见表1。

煤炭企业高管是领导者角色,建议由矿长牵头进行推动,实现强有力的支撑;信息化建设部门是核心的组织实施者,驱动整体过程的执行;数据治

理平台软件建设商是支撑者,为企业提供全面的服务;“采、掘、机、运、通”等业务部门是执行者,从业务角度使用系统,并提出需求;煤矿各子系统建设商是配合者,积极配合完成数据对接工作。

结 语

煤炭企业数据治理是长期、复杂的系统工程,呈现效果缓慢且持久,始于当下,利在千秋。建设好煤炭企业数据治理工程既需要在技术侧作出清晰的技术架构规划,又需要在业务侧明确规划和目标。系统建设效果的好坏不仅在于系统本身的投入,还在于管理制度的完善、人员的投入与培养,以及业务单位的配合和长期的坚持,只有相关责任单位各司其职,才能实现煤炭数据治理建设的成功。

■ 助理编辑:李雅楠

作者简介:

第一作者:韩培强,高级工程师,现任陕煤集团神木张家峁矿业有限公司机电技术管理中心高级业务主管。

E-mail: 420207504@qq.com

作者单位:陕煤集团神木张家峁矿业有限公司