

文章编号:1671-251X(2018)06-0016-05

DOI:10.13272/j.issn.1671-251x.2018010014

煤炭大数据平台建设的 key 技术及管理协同架构

谭章禄, 马营营, 袁慧

(中国矿业大学(北京)管理学院, 北京 100083)

摘要:分析了煤炭大数据平台应用的现状及需求;从细粒度、广范围的数据获取、灵活可扩展的数据存储、多模态的数据处理、专业的数据可视化、高质量的数据管理 5 个方面详细阐述了煤炭大数据平台建设的 key 技术,为煤炭大数据的采集、存储、分析、管理提供基础;提出了煤炭大数据管理协同架构,主要从技术和管理 2 个层面为煤炭大数据平台的建设提供参考,为大数据管理的顶层设计提供了思路。

关键词:煤炭工业;煤炭大数据平台;数据管理;数据可视化;协同管理

中图分类号:TD67 文献标志码:A

网络出版地址:<http://kns.cnki.net/kcms/detail/32.1627.TP.20180518.1616.002.html>

Key technologies and management collaborative architecture of construction of coal big data platform

TAN Zhanglu, MA Yingying, YUAN Hui

(School of Management, China University of Mining and Technology (Beijing), Beijing 100083, China)

Abstract: Application status and needs of coal big data platform were analyzed. Key technologies of construction of coal big data platform were explained in detail from five aspects of fine-grained and wide-range data acquisition, flexible and scalable data storage, multimodal data processing, professional data visualization and high-quality data management, the foundation for collection, storage, analysis and management of coal big data was provided. Collaborative architecture of coal big data management was put forward, suggestions for the construction of coal big data platform was provided from two aspects of technology and management, and ideas for top-level design of big data management was provided.

Key words: coal industry; coal big data platform; data management; data visualization; collaborative management

0 引言

煤炭大数据来源广泛,不仅包括智慧矿山的过程数据,还包括与经营相关的内外部数据,涵盖了煤炭企业生产经营整个生命周期中的所有数据及其发生的时间、空间关系^[1],这是煤炭行业亟需挖掘的“金矿”。鉴于煤炭行业的特殊性,煤炭领域的工业大数据规模将会是其他大数据领域的 2 倍^[2]。正是

由于生产环境及过程的复杂性,煤炭行业对数据的采集、处理和深度挖掘等多方面均存在盲点,煤炭企业正积极探索煤炭大数据的应用,并开始尝试构建煤炭大数据平台。

大数据平台是数据采集、存储、处理、分析等的综合载体,形成了统一的 IPO (Input-Process-Output,输入-处理-输出)模式,是大数据全生命周期发展进程的承载,为煤炭行业大数据深度挖掘提

收稿日期:2018-01-06;修回日期:2018-03-15;责任编辑:张强。

基金项目:国家自然科学基金项目(61471362)。

作者简介:谭章禄(1962—),男,江西赣州人,教授,博士,研究方向为信息化、可视化、大数据、信息标准化,E-mail:tanzl@vip.sina.com。通信作者:马营营(1992—),女,河北保定人,硕士研究生,研究方向为大数据、信息标准化,E-mail:yingying_jy@126.com。

引用格式:谭章禄,马营营,袁慧.煤炭大数据平台建设的 key 技术及管理协同架构[J].工矿自动化,2018,44(6):16-20.

TAN Zhanglu, MA Yingying, YUAN Hui. Key technologies and management collaborative architecture of construction of coal big data platform[J]. Industry and Mine Automation, 2018, 44(6): 16-20.

供了可靠支持。当前,我国煤炭大数据平台建设和应用一直处于探索阶段,且推进缓慢。本文在介绍煤炭大数据平台应用现状的基础上,分析了当前煤炭大数据平台建设需求,重点讨论了煤炭大数据平台建设的关键技术及管理协同架构。

1 煤炭大数据平台应用现状及需求

1.1 煤炭大数据平台应用现状

我国煤炭行业大数据平台应用还处于初级探索阶段,并存在一些问题。近几年,煤炭行业对煤炭大数据平台建设一直在进行着尝试和突破。2015年6月,国内首个煤炭大数据平台在太原上线,主要是运销数据的分析和应用^[3]。2016年7月,煤炭大数据平台V2.0升级版上线,增加了预警预测等功能^[4]。2017年12月,能源大数据平台正式启动,目前主要是集中交易信息,目标是为能源产业的宏观调控、生产调度、消费预测等方面提供支持^[5]。贵州省毕节地区设计应用了基于大数据的煤矿安全生产监管联网平台^[6]。部分煤炭企业也将大数据建设上升到战略层面。如有的煤矿加强实施大数据工程,把大数据应用提升至企业战略层面^[7];某智能矿山试点项目利用大数据中心进行数据挖掘,实现了生产态势感知与安全预警、资产状态管理与预防性维护等^[8],助力智能矿山建设;也有煤炭企业与互联网企业进行合作,计划进行数据和人工智能(Artificial Intelligence, AI)的深度应用。

我国煤炭大数据平台建设起步较晚,多数煤炭企业缺少对大数据平台技术和管理的全面认识,更多的大数据平台着重行业局部,如安全预警方面,而且矿井恶劣的生产环境导致数据采集比较困难,难以实现全矿井数据的整合,同时煤炭大数据标准的滞后也制约着煤炭大数据平台的建设。煤炭大数据平台的数据来源、数据整合体系和处理结果还有待完善,数据未能充分利用、交易渠道不畅通及未能挖掘大数据的隐含价值^[9]等问题有待解决。总的来说,煤炭大数据平台建设存在技术和管理2大层面的问题。本文通过介绍煤炭大数据平台建设的关键技术和管理协同架构,力图推动煤炭企业有效利用大数据。

1.2 煤炭大数据平台建设需求

基于上文对煤炭大数据平台的应用现状的分析,将煤炭大数据平台建设过程中的主要需求归纳为如下3点:

(1) 目标明确。企业或行业在煤炭大数据平台建设过程中,首先要了解大数据并确保大数据(数据

和技术)来源的可靠性,在此基础上,明确大数据平台在企业中的定位和建设目标是实现平台价值和企业利益的前提。目标导向不仅可以激励所有参与者排除障碍,在数据提取和采集、处理方面提供支持,而且可以为大数据平台建设提供持续动力和方向。

(2) 清晰的架构设计和技术选型。大数据技术在煤炭行业属于应用新兴技术,存在诸多不清晰的地方。煤炭大数据平台建设的网络环境搭建、硬件选型和采用的技术体系,可根据建设规模和目的来选定。同时大数据技术多种多样,按照怎样的要求来选择,需要选用哪些技术来达到目标等问题是煤炭大数据建设首先需要明确的内容。

(3) 科学的管理框架。煤炭大数据平台涉及范围广,涉及企业、组织和各类数据,对管理的要求十分高。谁来主导建设、怎么管理、如何确定管理内容、如何进行协调、如何完成顶层设计都是大数据平台建设亟需解决的管理问题。

2 煤炭大数据平台建设的关键技术

为实现大数据技术和煤炭工业大数据的最佳匹配,围绕大数据生命周期,根据煤炭大数据特性构建了煤炭大数据平台架构,如图1所示。煤炭大数据平台建设的关键技术主要包括数据获取技术、数据存储技术、数据处理技术、数据可视化技术,以及贯穿数据全生命周期的数据管理技术。

2.1 细粒度、广范围的数据获取技术

大数据分析的基础是数据的详细性、全面性、混合性,所以,煤炭大数据平台建设的首要目标是能够收集到细粒度、广范围的数据。细粒度数据是指数据颗粒足够小,比如瓦斯数据具有明确的时刻地点、前后数值,某工作面每小时平均进尺等,而不是统计的月度数据或年度数据;广范围数据是指矿井的机器、工作面、巷道和经营中的生产数据、环境数据、安全数据和管理数据等多维度的数据。

煤矿可用的数据获取手段多样,有常见的感知硬件,如RFID、摄像头、GPS、传感器等;有智能化矿用设备,如矿用机电设备、自动感知和手动写入并存的智能终端;有工程数字化模拟系统,如矿床数字模型、地表移动变形模拟系统;有感知控制系统,如矿井GIS系统、SCADA系统(Supervisory Control and Data Acquisition)、MES系统(Manufacturing Execution System)、矿井安全监控监测系统;有管理智能决策系统,如成本管控系统、企业资源计划(ERP)系统、OA系统、客户关系管理(CRM)系统、供应链管理(SCM)系统、业务流程管理(BPM)系统

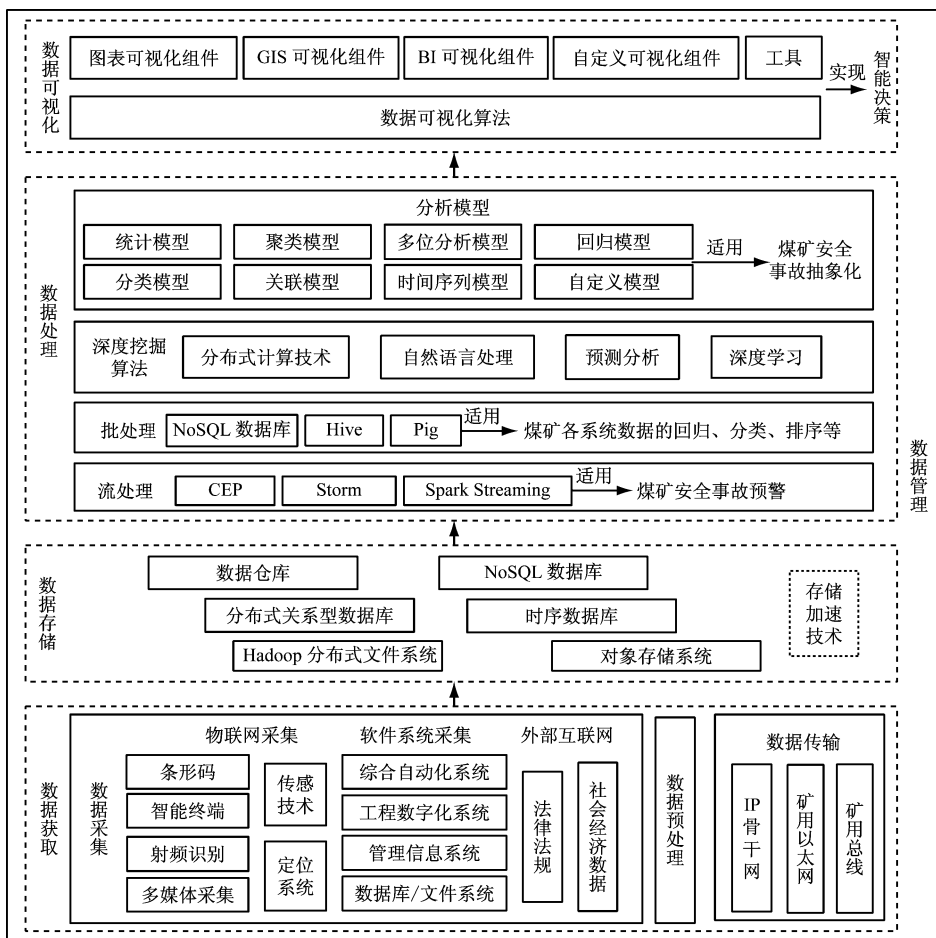


图 1 煤炭大数据平台技术架构

Fig. 1 Technology architecture of coal big data platform

等。还包括通过外部网络获取技术,如与外部系统数据交换、flume 等互联网数据爬取技术、第三方数据服务访问。获取的数据类型也多样,如实时性数据、离线批量数据、结构化数据、异构化数据等。通过小颗粒、广范围、规范化的数据采集可实现从智能机器、智能生产模式、资源协同到服务创新过程的互联互通。

数据采集后,需进行数据预处理,即对不规则、不准确的数据集进行反复的 ETL (Extract-Transform-Load,抽取-转换-加载)处理,可以结合统计学方法或 SQL 等多种方法,如采用 VAR 模型处理异常数据、用 SQL 挑拣错误数据、规则性粒度聚合。结合煤矿企业的实际条件和要求,将采集的数据进行初步(向外)传输。

2.2 灵活可扩展的数据存储技术

集合时间、空间的大量多源异构数据要求数据存储框架具有可扩展性和容错能力,并需要高效的存储管理方法。对于大数据存储框架,可应用的平台较多。Hadoop 分布式文件系统、基于分布式文件系统的 NoSQL 数据库、分布式数据库 HBase、

GoodData、关系数据库、对象存储系统和时序数据库等都适合捕捉和存储海量数据^[10]。针对煤炭不同场景产生的结构化业务数据、时序的设备监测数据和非结构化数据,一方面可通过分布式文件系统、对象存储系统和时序数据库等多种存储框架进行综合运用,另一方面可通过统一规范定义企业数据的核心语义结构,实现对数据的跨界连接、组织融合和协同查询。

2.3 多模态的数据处理技术

大规模 GPU (Graphics Processing Unit,图形处理器)集群并行计算的分布式计算技术、自然语言的处理、预测分析、基于多层神经网络的深度学习等先进分析技术都可用于煤炭大数据处理中,但需要做特定的模型数据匹配学习,比如煤矿的安全指数由瓦斯等气体、顶板压力、矿压等指标构成,复杂的模型更需要大量历史数据对模型进行训练学习,以便更深入地挖掘其应用价值。

大数据可分为实时性数据和非实时性数据。煤炭大数据应根据不同的应用场景运用不同的分析处理技术,如批处理、迭代处理、流处理、交互式分析

等,进而充分利用大规模数据^[11]。一方面,煤炭企业可通过分布式文件系统(如基于 NoSQL 数据库或 Hive、Pig、Spark、AdaBoost 等)实现离线批量处理或迭代处理,比如,对各个信息系统产生的海量生产或过程数据进行回归、分类、排序、协同过滤等。另一方面,可通过 Storm、Spark Streaming 等实现数据实时在线处理,比如,煤炭安全事故预警、矿井环境变化分析等。

2.4 专业的数据可视化技术

数据展示又称数据可视化,其目的是更好地解释数据分析结果,以使用户获取信息价值。煤炭大数据可通过 ECharts、ggplot2 等做图软件及诸如 Tableau、FineBI 的新型自助式 BI (Business Intelligence, 商业智能) 工具等可视化工具,以常规图形——折线图、饼图、散点图、气泡图、树形图或箱型图等展现形式在移动终端、PC 和各类信息发布系统上呈现出优秀的解释型可视化作品,进而用户通过操作界面与数据交互实现进一步的分析和利用。李光达^[12]针对煤炭行业数据展示方式进行了详细的分析。展示方式的选择应考虑使用人群,煤炭大数据交易面向的人群不同,数据的展示方式也要求不同,比如研究人员需要原始数据或结果数据集,而企业或行业领导只需要直观的图表。

2.5 高质量的数据管理技术

数据管理覆盖煤炭大数据全生命周期过程,需验证在生命周期中的每个过程是否都能够被大数据系统正确处理^[13]。大数据产品(被管理对象)不仅量上有极大的提升,而且更多需要关注管理节点之间的调度、调优和监控,需要考虑数据仓库相关的数据管理技术^[14]。同时,随着管理节点海量增加,技术要不断创新,需实现数据管理能力的移动化。煤炭大数据可通过 Hortonworks 管理架构实现基于角色的用户认证、授权和审计功能,通过 Cloudera 管理架构实现对多节点的集中安装部署、监控预警等核心的管理功能。

3 煤炭大数据管理协同架构

煤炭大数据具有复杂多样的特性,尤其是安全数据、地质数据和采掘数据具有较强的专业性,处理分析困难,管理难度更是极高,仅有技术体系不能使大数据平台科学有序地运行并提供服务。所以,需要专业的管理人员、知识发现研究人员、系统协调人员和技术人员共同协作,进行数据的收集、处理、挖掘、展示及数据共享和交易,为此,提出了煤炭大数据管理协同架构,如图 2 所示。该架构对煤炭大数

据协同管理进行了清晰的展示,可为顶层设计提供参考。

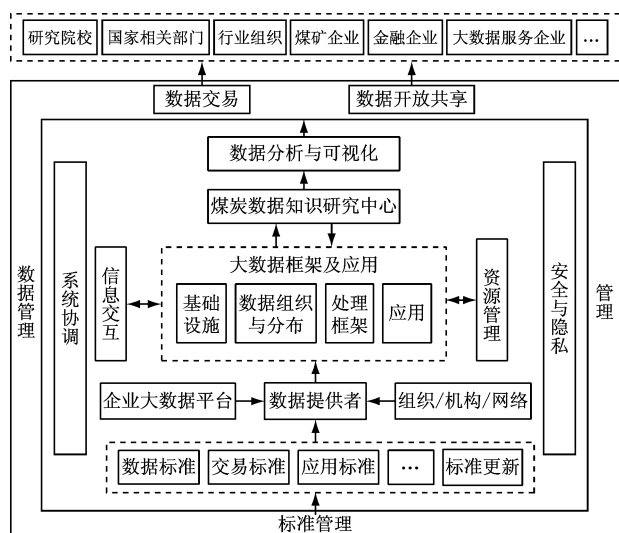


图 2 煤炭大数据管理协同架构

Fig. 2 Collaborative architecture of coal big data management

该架构简化地表示了贯穿大数据全生命周期的技术框架和大数据实际应用的各角色,重在体现多方资源的协同管理关系,以数据的流动为框架主线(用箭头表示),所有的数据处理和流向均处于管理的大环之中;垂直轴向是数据流向,水平轴向是数据流动的保障因素。

架构图中的外框是管理层面,该层面与内部的数据流动和协调均存在影响和交互作用,主要包含大数据生命周期的数据管理、标准管理、大数据交易管理及统一协调管理。标准管理是大数据管理的先行工作,是做好每个环节紧密耦合的保障工作;大数据交易管理是数据流动的最后一个环节,是系统与外界的知识交流窗口,对大数据的采集和分析具有一定的指导作用;数据管理伴随着大数据生命周期的每一个过程,贯穿该架构内部的每一个环节。除此之外,管理还包括调配、配置、系统运行情况、协调者、研究中心等的统一协调管理,只有每个环节都管理协调好,煤炭大数据平台才能有效运行。

架构图中的内框涵盖了围绕数据流动产生的平台中的各角色、功能、组件及相互之间的作用关系,即为管理的对象,其中系统协调和安全管理与隐私管理是大数据平台管理的关键要素。系统协调功能是通过信息架构师、软件架构师等各种角色共同完成的,每个角色互相配合操作使得整个系统无障碍运行;安全与隐私管理需构建全方位的安全防护体系、预警机制,提升容灾能力,以保障从数据收集到数据交易的正常进行。

数据提供者可以是煤炭企业的大数据平台,也可以是加入行业大数据平台的组织、机构或互联网,数据提供者还需要一套保障机制;资源管理和信息交互规范为形成一个开放、共享的煤炭大数据提供了资源基础和接口基础;培养煤炭大数据分析挖掘等方面的专业人才是将数据知识化的关键。

此外,该框架中涉及各主体(数据提供者、系统协调者、标准化工作者、技术架构提供者、数据管理者、平台管理者)的协作模式,可以有多种形式,比如行政模式、市场模式、联盟模式等。行政模式指政府主导的管理模式,能为平台提供坚实的后援保障,利于平台建设成功;市场模式指市场化、以营利为目的的模式,能吸引更多的组织加入,利于平台运行平稳;联盟模式指大型企业或组织以共同发展为目的的模式,有利于数据保障和结果应用检验。单个协作模式或组合模式都各有利弊,哪种更利于大数据平台建设发展还应结合博弈论等方法 and 实际情况具体分析。

4 结论

(1) 大数据平台化是煤炭大数据发展的必然方向,在大数据平台建设中,物联网和智慧矿山的建设对保障数据采集的详细性发挥了关键作用;大数据平台的技术体系或技术框架的选择要考虑煤炭工业的特性,比如安全领域的实时性、矿井环境的复杂性、地质模型的特殊性、数据的异构性等,根据具体建设目标和规模来选择;必要时可采用存储加速技术、资源管理技术等辅以处理分析,数据管理质量直接关系到平台的成功利用。

(2) 不论技术选型如何,大数据的管理协同思想都必须时刻贯穿在建设过程中,本文提出的大数据管理协同架构为大数据管理的顶层设计提供了思路。

(3) 云存储和云计算发展逐渐成熟,煤炭大数据部署到云系统上,将大数据服务云化,为煤炭大数据平台成功实施提供了技术保障。

参考文献(References):

[1] 谭章禄,马营营. 煤炭大数据研究及发展方向[J]. 工矿自动化,2018,44(3):49-52.
TAN Zhanglu, MA Yingying. Research on coal big data and its developing direction [J]. Industry and Mine Automation, 2018,44(3):49-52.

[2] 周兴社. 工业大数据特点、价值及其计算[EB/OL]. (2017-10-13) [2017-11-26]. <https://wenku.baidu.com/view/e68d87d6e518964bce847c0e.html>.

[3] 邢云鹏. 中国煤炭大数据平台上线[N]. 中国国土资

源报,2015-06-20(002).

[4] 中国煤炭市场网. 煤炭大数据平台 V2.0 全新上线[EB/OL]. (2016-07-19) [2017-11-26]. <http://www.cctd.com.cn/show-416-144772-1.html>.

[5] 新华网. 中国(太原)煤炭交易中心启动能源大数据平台[EB/OL]. (2017-12-06) [2017-12-26]. http://www.sx.xinhuanet.com/2017-12/06/c_1122069303.html.

[6] 沈宇,王祺. 基于大数据的煤矿安全监管联网平台设计与实现[J]. 矿业安全与环保,2016,43(6):21-24.
SHEN Yu, WANG Qi. Design and implementation of coal mine safety supervision networking platform based on big data [J]. Mining Safety & Environmental Protection, 2016,43(6):21-24.

[7] 杨林. 兖矿集团煤矿工业大数据发展应用刍议[C]//第26届全国煤矿自动化与信息化学术会议暨第7届中国煤矿信息化与自动化高层论坛,济南,2017.

[8] 韩建国. 神华智能矿山建设关键技术研发与示范[J]. 煤炭学报,2016,41(12):3181-3189.
HAN Jianguo. Key technology research and demonstration of intelligent mines in Shenhua Group [J]. Journal of China Coal Society, 2016,41(12):3181-3189.

[9] 马小平,代伟. 大数据技术在煤炭工业中的研究现状与应用展望[J]. 工矿自动化,2018,44(1):50-54.
MA Xiaoping, DAI Wei. Research status and application prospect of big data technology in coal industry[J]. Industry and Mine Automation, 2018,44(1):50-54.

[10] 于富东. 大数据平台的关键技术及组网方案[J]. 电信科学,2015,31(7):164-169.
YU Fudong. Key technologies and networking scheme of big data platform [J]. Telecommunications Science, 2015,31(7):164-169.

[11] 廖建新. 大数据技术的应用现状与展望[J]. 电信科学,2015,31(7):7-18.
LIAO Jianxin. Application status and prospects of big data technology [J]. Telecommunications Science, 2015,31(7):7-18.

[12] 李光达. 基于认知科学的煤矿安全可视化管理效应研究[D]. 北京:中国矿业大学(北京),2017.

[13] 全国信息技术标准化技术委员会大数据标准工作组,中国电子技术标准化研究院. 大数据标准化白皮书(2016版)[R/OL]. (2016-07-26) [2017-11-28]. <https://wenku.baidu.com/view/1b3b63c03169a-4517623a34a.html>.

[14] 孟小峰,慈祥. 大数据管理:概念、技术与挑战[J]. 计算机研究与发展,2013,50(1):146-169.
MENG Xiaofeng, CI Xiang. Big data management: concepts, techniques and challenges [J]. Computer Research and Development, 2013,50(1):146-169.