

数据湖平台智能油田实时数据服务标准化研究

牛永胜¹, 岳翔², 陈维汉¹, 王思楠³

(1. 中海石油(中国)有限公司天津分公司, 天津 300450;
2. 中海油研究总院有限责任公司, 北京 100028; 3. 中国电子系统技术有限公司, 北京 100089)

摘要: 智能油田应用开发涉及的数据种类繁多、来源多样, 数据量大且标准不一。勘探开发数据湖平台的目标是在数据源与数据应用之间建立桥梁, 对采集到的原始数据在数据湖内进行标准化治理、聚合加工并封装为标准化服务提供应用系统调用, 实现数据资源的管用分离。根据某智能油田项目所开展的实时数据服务入湖迁移与标准化改造工作, 提出了一套基于数据湖平台的智能油田实时数据服务标准化开发方法, 并利用数据湖平台实时数据开发工具, 将智能油田实时数据服务经标准化改造后迁移入湖。该项研究为油气生产企业数智化转型提供了新的方法与工具支撑, 有助于提高智能油田大数据分析应用的开发效率与质量。

关键词: 数据湖; 智能油田; 实时数据服务; 标准化开发; 数据治理

中图分类号: TE19; TP18

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2024.09.013

引用格式: 牛永胜, 岳翔, 陈维汉, 等. 数据湖平台智能油田实时数据服务标准化研究 [J]. 网络安全与数据治理, 2024, 43(9): 77-83.

Standardized development and practical implementation of real-time data services for intelligent oilfields based on a data lake platform

Niu Yongsheng¹, Yue Xiang², Chen Weihang¹, Wang Sinan³

(1. CNOOC (China) Tianjin Branch, Tianjin 300450, China; 2. CNOOC Research Institute, Beijing 100028, China;
3. China Electronics System Technology Co., Beijing 100089, China)

Abstract: Intelligent oilfield application development involves a wide variety of data with large amount, from diverse sources and different standards. The Exploration and Production (E&P) data lake platform is positioned to build a bridge between data sources and data applications. The collected raw data is standardized, aggregated and processed in the data lake and encapsulated into standardized services for application systems to call, thereby realizing the separation from data resources to data applications. This paper summarizes the work related to the migration of real-time data services into data lake and the standardized transformation during one intelligent oilfield project. Then a standardized approach for real-time data services development in intelligent oilfield based on the data lake platform is proposed, which has been implemented and verified by using data lake real-time development tool to standardize the real-time data services and then migrate into data lake. This research provides new method and tool support for the digital and intelligent transformation of oil and gas enterprises to improve the efficiency and quality of industrial big data application development.

Key words: data lake; intelligent oilfield; real-time data service; standardized development; data governance

0 引言

目前, 大多数油气田企业仍然使用传统的以专业为单位“竖井”式数据管理手段。随着多源异构数据数量和种类不断增多, 油气田生产涵盖了从地质勘察、物探管理、油气集输管理、设备设施状态监测到健康和安全管理等流程和工艺。每个环节产生的数据都各具特点,

导致数据管理分散凌乱, 给数据资源的上下衔接和循环共享管理带来困难。随着大数据时代的演进, 数据管理手段不断更新, 数据湖被广泛接受。作为一种数据资源治理与服务管理平台, 数据湖接收来自不同源头、格式和结构各异的数据, 并对数据进行清洗、转化与加工处理, 封装为数据服务, 供各应用系统使用。

随着数字孪生 (Digital Twins, DT)、商业智能 (Business Intelligence, BI)、模型诊断等技术的发展,油气田生产的业务侧对实时数据分析应用的需求日益迫切。石油行业在利用实时数据服务时又面临着几个显著的挑战:一是数据量庞大且种类繁多,要求石油企业必须具备高效的数据采集、存储和处理能力,以确保数据的完整性和可用性。二是实时性要求极高,因为石油生产过程中的任何细微变化都可能对生产效率和安全生产产生重大影响,因此实时数据服务必须能够迅速捕捉并响应这些变化。此外,数据安全与隐私保护也是石油行业必须面对的重要问题。

智能油田实时数据服务主要通过边缘计算与云端协同、大数据分析可视化、智能决策支持以及一体化管理平台等方法实现。边缘计算与云端协同在实时性和数据处理能力上表现突出,能够显著降低数据传输延迟并提升处理效率;智能决策支持系统则显著增强了决策的科学性和准确性。而大数据分析能够为油田的勘探开发、生产优化、设备维护、风险预警等提供支持。作为实时数据应用的基础,面向大数据分析的实时数据服务标准化开发是一项决定性技术,保证实时数据能够在数据湖中实现采集、加工、共享、存储、应用,保证整个数据链路完整、准确,保障数据调用质量和数据统一^[1-4]。现有的实时数据标准化在数据采集、处理和存储方面面临复杂挑战,本文提出了一种基于分层结构的实时数据服务标准化方法。该方法不仅精细化了数据分层,完善了管理体系,还提供了切实可行的解决方案,有效应对了智能油田实时数据服务标准化中的各种难题。

1 智能油田实时数据服务现状分析

1.1 实时数据服务存在的问题

目前,勘探开发数据湖已经基本实现勘探、开发、生产、工程、钻完井等业务数据的统一采集、传输、存储和应用。随着智能油田建设的深入推进,对数据服务提出了新的需求,传统结构化数据服务模式不足以支撑如今的智能油田发展需求^[5-11],主要体现在以下几个方面:

(1) 数据服务的质量与安全性难以保证。处理和管理庞大、多样化的数据集时,维护数据的准确性和完整性变得极为复杂。特别是在物联网平台这样的动态环境中,由于数据源的多样性和不断增长的数据量,实现数据的实时处理和访问,同时确保数据质量和安全性,成为了一项艰巨的任务。实时流动性和可访问性管理对于防止数据滥用或泄露至关重要,但这在高速和大规模数据处理环境中尤为困难。

(2) 难以实时结合增量数据和全局性数据。在提供实时数据服务时,如何高效地结合实时计算与增量数据处理,并与全局性数据聚合和分析策略相协调是主要挑战之一。这要求技术解决方案能够灵活处理实时数据流,并与全局性的数据汇总策略无缝对接,以支持快速响应的业务需求和高效的数据分析。

(3) 数据一致性和实时复杂分析难以保证。在实时数据服务中,维持数据一致性并进行实时复杂分析是一项重大挑战。由于缺乏有效的数据预处理和结构化数据,实时服务中的数据分析因此变得困难。这些挑战反映了需要在确保数据一致性的同时,提供对复杂数据分析的支持。

1.2 实时数据服务标准化发展目标与挑战

数据服务标准化是一种对数据进行统一规范、减少差异、消除冗余的过程,它能够提高数据的质量和可用性,降低数据管理的成本和风险。实时数据服务标准化旨在确保实时数据在不同环节的采集、传输、处理和应用中达到一致性和高质量,提高数据服务管理的效率和可靠性^[12]。智能油田实时数据服务标准化面临着各种复杂而紧迫的挑战^[13-14],为行业创造一个统一、高效、安全的数字化工作环境,推动智能油田实时数据服务管理水平的全面提升^[15-17]。

(1) 数据采集面临实时数据来源众多且异构的挑战,制定一致的实时数据服务结构标准变得尤为复杂。数据采集标准化目标为保证所采集数据具有准确性、完整性和及时性。

(2) 数据处理面临挑业务服务不规范、数据源的结构需整合,确保实时数据的一致性和互操作性的挑战。数据采集标准化目标为确保数据在进行清洗、去重、格式转换等操作时的可用性和可操作性。

(3) 数据存储面临传统数据库扩展遭遇瓶颈、海量数据接入难,业务之间耦合高,难以完成数据切分。数据采集标准化目标为选择合适的数据存储标准化方案,建立完善的数据备份和恢复机制,确保实时数据的可靠性和安全性。

综上,推出普适性强的标准化改造方案来解决这些难题迫在眉睫。本文从实时数据服务结构和入湖两个方向实施标准化开发,力求取得全面且切实的效果。

2 智能油田实时数据服务标准化开发方法

以油田生产业务域为例,在生产过程需要采集和处理大量的数据,如油气田生产动态数据、设备设施运行维护数据和质量健康安全环保 (Quality Health Safety Environmental, QHSE) 数据等。

智能油田前期的实时数据服务基于列式存储数据库，在海量实时数据写入的情况下，数据延迟等问题较为明显。同时，数据治理的整体建设尚不完善，数据分层分域规划不够清晰，可能导致数据使用和维护困难、数据重复计算和开发、以及数据层级混乱等问题。为了解决这些问题，本文提出了一种实时数据服务标准化方法，建立了包括贴源层（Operational Data Store，ODS）、明细层（Data Warehouse Detail，DWD）、汇总层（Data Warehouse Summary，DWS）、维度层（Dimension，DIM）和应用层（Application Data Service，ADS）在内的五层结构实时数据服务管理体系，如图 1 所示。此方法不仅完善了数据服务管理体系，使数据分层更加精细化，还解决了智能油田实时数据服务标准化过程中遇到的数据采集、数据处理和数据存储等挑战，提供了应对实时数据服务标准化问题的切实可行的解决方案。

2.1 数据采集标准化

为了实现保证所采集的多源异构数据具有准确性、完整性和及时性，需要实现数据采集标准化。因此，在实时数据服务管理体系，本文通过建立贴源层（ODS）令原始多源异构数据存储得到保障，使系统做到既算也存，实现增量数据和全局性数据相结合。通过在 ODS 层中开辟实时多源异构数据集定义标准化模式，采用屏蔽底层复杂业务策略，将多源异构数据暴露给应用层，将底层业务变动与上层需求变动对模型的影响降至最低，使业务变化影响主要集中在基础数据层。同时，结合自上而下的建设方法，进一步削弱需求变动对模型的影响。在数据层面，以智能油田实时数据业务需求为导向，依托勘探开发数据湖平台实时数据服务管理体系规划，设计实时数据服务管理体系。

对于智能油田实时数据及服务标准建设，制定简洁明确且具有描述性的命名规范；明确数据采集、处理、存储、服务等各个环节的业务流程，确保各环节无缝衔接的业务规范；依据业务需求和数据特点，设计合理的数据模型，建立数据开发规范；通过服务接口设计、服务性能优化以及安全性管理进行数据服务开发和应用规范。经由需求分析、标准制定、体系设计、开发实施、测试验证与优化改进等实施步骤，最终形成实时数据及服务开发范式。

2.2 数据资源处理标准化

在处理异构数据资源的过程中，业务服务类型与数据采集和处理过程密切相关。由于业务需求的多样性，重用现有的数据处理流程可能会变得困难。业务环境的变化和复杂性进一步增加了难度，即使在调研过程中收集了完整的数据资源信息，如果没有对异构数据资源进行标准化处理，也很难保证数据可用性与可操作性。由不同单位或部门提出适应独特需求的数据资源处理标准必然会存在差异，难以适用于整个项目的数据处理流程。

本文提出了一种明细数据的层级（DWD）来进行数据处理过程中标准化改造。具体操作流程为：首先 DWD 采用 Flink 流计算，完成 ODS 中的数据实时计算、补点、过滤、清洗、转化、关联等操作，形成针对不同业务主题的 DWD 数据明细层，使得数据质量更高，字段更全面。其次，DWD 在初步完成数据流处理后，会将明细数据全量入湖，为实时数仓提供明细数据。最后，以实时数据服务总体数据结构进行实例化分析，数据资源处理以业务服务主题为导向，保持业务架构、逻辑一致性。

具体实施过程为，DWD 在业务域的划分上采用了“业务域 - 一级业务 - 业务活动 - 业务手段/主题”四层

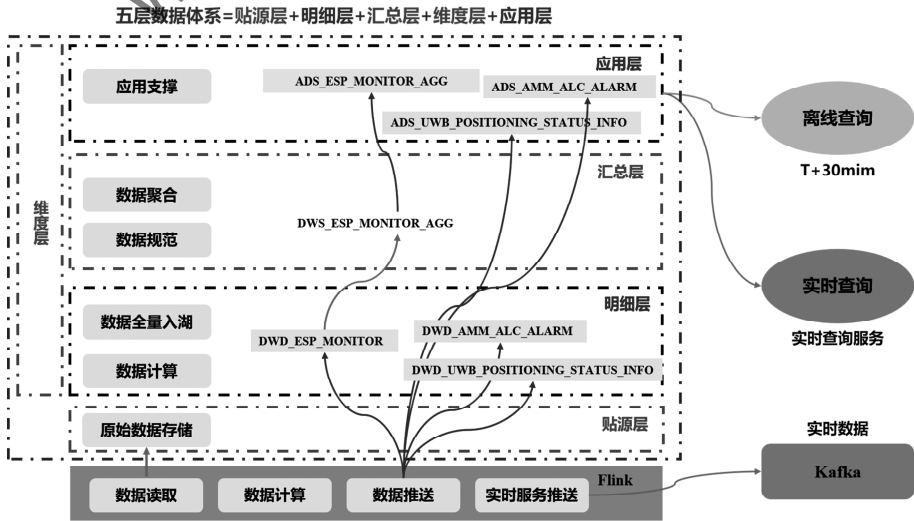


图 1 实时数据服务管理体系

结构。一方面能够较为精确、系统地描述油田企业的各种业务活动;另一方面又不使层级过于复杂,增加数据架构节点和难度。该体系架构将油田企业的业务划分为不同的业务域,每个业务域又细分为不同层级的业务。例如,可以将业务域划分为井筒工程、综合研究、海洋与地面工程、油气田生产等业务域。每个业务域中又包含着不同的层级业务。例如在油气田生产业务域包含了一级业务:油气田生产动态、设备设施运行维护和QHSE(质量健康安全环保)。在低层业务中,涉及具体的各种业务手段/主题,例如:采出井状态管理、油气处理单元集输管理和电气设备监测等。通过制定这种数据业务域标准化规范,使得油田企业的业务结构和信息框架变得更加清晰。同时,该体系架构提供了更加便捷的方式来描述和整理企业的各种业务活动,更好地支持企业的业务管理和信息管理。

2.3 数据服务存储标准化

本文通过在实时数据服务管理体系中建立维度层(DIM)来解决数据存储标准化问题,确保实时数据的可靠性和安全性。DIM用于存储聚合规则表、维度表以及点位信息注册表等维度数据,能够建立一致数据分析维表,可以降低数据计算口径和算法不统一风险,保证了数据一致性,并使下游复杂数据分析实施效率得到提升。该模块包括智能油田实时数据服务标准化开发入湖操作,入湖过程致力于实现对实时标准化数据服务的存储管理。图2显示的是实时数据入湖及服务开发架构,一方面在勘探开发数据湖底座搭建高可用的Kafka集群、Flink计算引擎集群和MPP实时数仓,作为智能油田实时数据的计算和存储支撑;另一方面数据湖平台提供实时数据开发管理和数据服务管理体系,使开发者能够更便捷地开发和

集成智能油田数据服务。查询服务通过数据湖服务管理平台进行服务开发,向业务模块提供rest API数据查询服务接口。实时数据推送服务依托Kafka,向业务模块提供数据推送服务。数据湖平台依托高性能的数据库和存储设备性能,实现数据存储和备份。此外,实时数据入湖及服务开发架构替代原有离散组件搭建方式,实现实时数据采、存、管、用一体化。

2.4 智能油田面向对外业务开发管理模式

针对对外业务开发,设计了汇总层(DWS)和应用层(ADS)。其中DWS内添加标准化数据聚合和数据规范流程,可以依据业务需求,面向主题来组织数据,给出数据组织规则,减轻分析查询压力,提升服务规范水平,确保了数据服务的质量与安全性。ADS提供面向服务的数据支撑,包括:(1)提供实时明细查询:Kafka(或ODS层)的数据使用Flink的实时计算进行过滤、清洗、转化、关联等操作,形成针对不同业务主题的主题表,针对有实时明细查询的业务需求,主题表直接写入ADS层,对外提供数据。同时,该明细表又作为其他应用的明细表同步写入DWD用于全量入湖及离线聚合计算。(2)提供离线聚合查询:实时数据高度汇总,面向业务主题,提供应用数据支撑。

2.5 预期成效

本文提出的实时数据服务标准化方法通过精细化的数据分层,有效解决了数据采集、处理和存储三个方面的实时数据标准化问题。主要优势如下:首先,它使数据结构清晰化,每一层级能够高效完成特定功能,从而简化后续的维护和治理。其次,在面对复杂问题时,数据分层能够实现快速任务拆分和工作分配。此外,规范化的数据分层显著减少了重复的数据计算和开发工作,

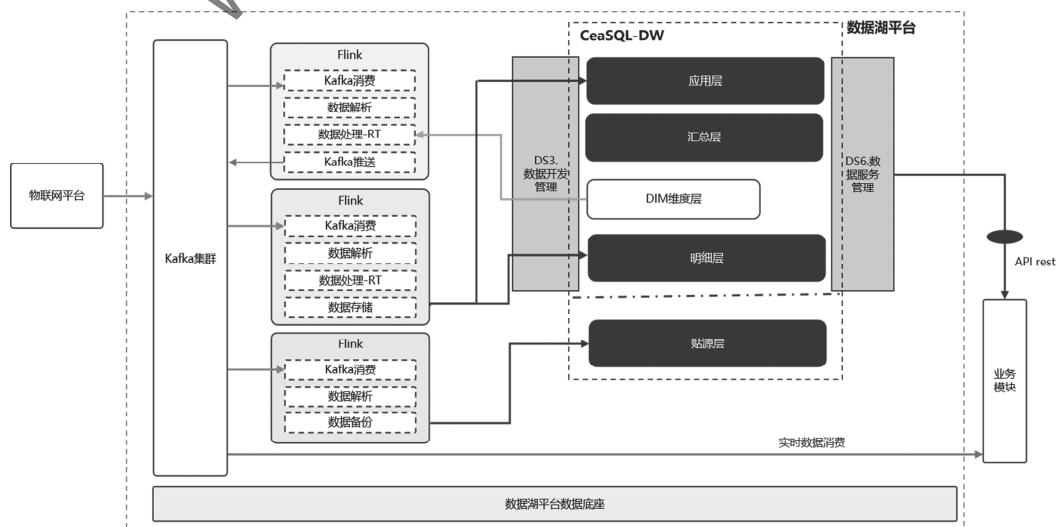


图2 实时数据入湖及服务开发架构

降低了数据管理成本。通过实施上述实时数据服务标准化开发方案，可以克服油气田智能化转型过程中存在的诸多难题，预期成效体现在两个方面：

(1) 预期功能型目标

一是形成基于数据湖平台智能油田实时数据服务标准化开发方法：基于集团的勘探开发数据湖平台，完成了实时数据服务开发和建设的标准化以及数据入湖过程的标准化设计，并在智能油田项目中成功转化成果。此过程确保了采集数据的准确性、完整性和及时性。数据经过清洗、去重、格式转换等操作，保证了其可用性和可操作性。通过选择合适的数据存储标准化方案，建立了完善的数据备份和恢复机制，确保了实时数据的可靠性和安全性，有效地克服了油气田智能化转型存在的难题，支撑了智能油田实时数据业务需求。

二是改善与应对智能油田数据管理问题：智能油田实时数据服务标准化开发为智能油田的油气田勘探、开发、生产、经营等全业务流程提供了标准的规范，准确指定业务域划分规则，保障了系统中数据服务的质量与安全性；所提出的标准化方法完善了传统数据湖的数据服务管理体系，极大地加强了上层服务的功能，改善增量数据和全局性数据结合效果；部门统一实施的数据服务标准化管理方法，将数据资源按照一套清晰且规范的标准化数据服务管理方法进行，有效解决数据一致性和实时复杂分析困难问题。

(2) 预期非功能型目标

全面提升智能油田建设项目智能化管理水平：通过结合标准化数据治理手段，提升了油田管理系统的功能，完善了数据服务管理制度，并加强了智能油田管理系统的建设。所提出的实时数据服务标准化方法，使数据采集、存储和服务更加高效，从而提升了智能油田系统的整体运行效率。此外，标准化数据服务管理制度确保了数据的质量和可靠性，为智能油田的全面智能化发展和推广奠定基础。

3 智能油田实时数据服务标准化开发实践

智能油田实时数据服务标准化方法可作为一种通用的数据治理手段，可广泛适用于数据量过剩的智能油田应用系统项目中。具体实践案例为：在某智能油田建设项目，实时数据服务标准化方法已正式完成项目落地。基于某勘探开发数据湖平台完成贴源层（ODS）、明细层（DWD）、汇总层（DWS）、维度层（DIM）和应用层（ADS）的五层结构实时数据服务管理体系的建立。实时数据服务管理体系会将实际接收数据源按照各自的标准化守则进行实时数据标准化处理，并分层全量入湖。图3所展示的是在某勘探开发数据湖平台中数据开发管理界面，该界面可视化了ADS、DWD、DIM、ODS、DWS具体于数据湖中所在的位置。这五层结构的实时数据服务管理体系会对数据实现实时开发管理和离线开发管理，可见所设计的标准化改造方案令数据源结构清晰透彻，便于工程师对数据资源头梳理以及对数据信息的维护。

此外，数据湖平台内置的API监控工具能够实现数据服务运行过程的实时监控。图4展示了2024年6-7月的数据湖内数据服务API调用情况和API的自验证界面，通过可视化服务调用情况进行测试，表明平台能够为工程师的数据运维提供稳定性和可靠性支持。以2024年6月25日当日调用情况为例，当日调用次数为26575次，调用失败率为0，表明平台能够为工程师的数据运维提供稳定和可靠的支持。图5展示的细节代码表明，该工具调用的数据服务API接口与工程师的调用结果一致，能够完成平台指令，这进一步确保了数据湖平台的稳定运行。

综上所述，基于数据湖平台的智能油田实时数据服务标准化开发已在某智能油田建设项目中成功实施，证实了实时数据服务开发建设标准化和入湖过程标准化方法的可行性。

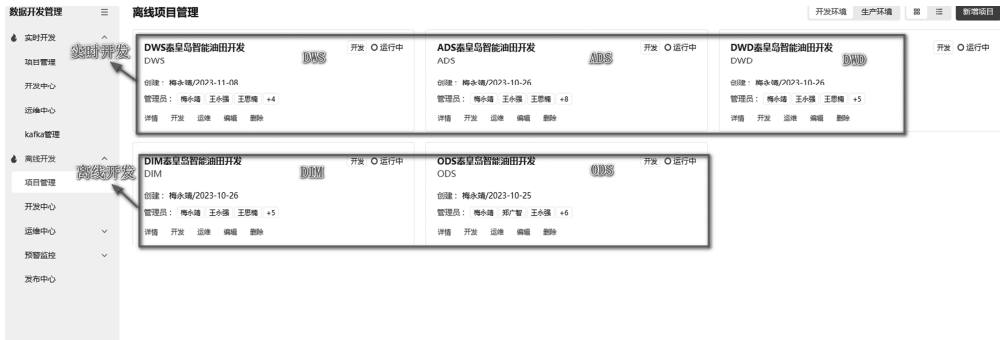


图3 数据服务项目管理



图4 数据服务 API 自验证



图5 第三方 API 接口功能性验证

4 结论

本文提供了一套完善的智能油田实时数据服务标准化开发方法,优化了原始数据标准化治理、聚合加工并封装等管理流程,为系统调用供应标准数据服务,较好地解决了数据服务质量与安全性难以保证、增量数据和全局性数据难以完美结合、数据一致性和实时复杂分析难以确保等问题。此外,搭建了勘探开发数据湖平台,并在智能油田建设项目中进行了实践,验证了标准化改造方法的可行性。在数据湖内,开发了多种模式的数据服务调用监控手段,实现了数据管理、使用和监控的闭环,为油气田智能化升级做出了积极贡献,并为未来基于数据湖平台的智能油田升级创造了条件。

参考文献

[1] 杨勇,黄文俊,王铁成,等. 梦想云数据连环湖建设研究[J]. 中国石油勘探,2020,25(5):82-88.
[2] 张芸. 浅谈石油勘探行业数据湖建设中的数据治理问题[J]. 中国管理信息化,2021,24(9):122-124.
[3] 尚新路,肖楠,严峻,等. 塔里木油田区域湖数据治理[C]//2023国际石油石化技术会议论文集,2023.

[4] HLUPIĆ T, OREŠČANIN D, BARANOVIĆ M, et al. An overview of current data lake architecture models [C]//2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), 2022.
[5] 王颖,顾娟,杨勇. 大型油田数据治理工程实践[J]. 中国管理信息化,2023,26(3):162-167.
[6] 李瑛,颜廷龙. 航空大数据的融合及挖掘技术综述[J]. 航空计算技术,2020,50(6):124-128.
[7] 邱利瑞,徐震,王者云,等. 面向智能油气田的数字化管理体系建设与实践[J]. 国际石油经济,2022,30(7):53-63.
[8] 徐鹏,高健伟,陈溯,等. 勘探开发数据资产化管理实践与思考[J]. 石油科技论坛,2020,39(5):34-40.
[9] 郭以东,马建国,余洋,等. 油气田企业能效数据融合研究与实践[J]. 中国石油勘探,2021,26(5):38-48.
[10] 王同良. 油气行业数字化转型实践与思考[J]. 石油科技论坛,2020,39(1):29-33.
[11] 符强,谭忠健,郭明宇,等. 渤海油田勘探录测试专业数据治理实践与思考[C]//2023中国油气人工智能科技大会论文集,2023.
[12] 刘曙光. 勘探开发数据标准化治理方法探讨[J]. 江汉石

- 油职工大学学报, 2021, 34 (5): 62-64.
- [13] 杜金虎, 时付更, 杨剑锋, 等. 中国石油上游业务信息化建设总体蓝图 [J]. 中国石油勘探, 2020, 25 (5): 1-8.
- [14] 石玉江. 智能油田在中国的研究现状分析 [J]. 海峡科技与产业, 2016 (12): 81-83.
- [15] 杜金虎, 时付更, 张仲宏, 等. 中国石油勘探开发梦想云研究与实践 [J]. 中国石油勘探, 2020, 25 (1): 58-66.
- [16] 万广峰. 巴西深水油气勘探开发实践 [M]. 北京: 石油工业出版社, 2020: 53-55.

- [17] 杜金虎, 杨剑锋, 张仲宏, 等. 中国石油勘探开发梦想云研究与应用 [M]. 北京: 石油工业出版社, 2020.
- (收稿日期: 2024-07-01)

作者简介:

牛永胜 (1984-), 男, 在职研究生, 工程师, 主要研究方向: 海上智能油田建设与管理。

岳翔 (1988-), 男, 硕士, 高级工程师, 主要研究方向: 人工智能、机器学习在油气行业的应用、勘探开发数据服务。

陈维汉 (1982-), 男, 本科, 工程师, 主要研究方向: 能源行业自动化、业务数字化、物联网应用。

(上接第 76 页)

- [27] 许翠霞. 动物真的能够成为法律主体吗? ——关于法律主体的前提性说明 [J]. 安徽大学学报 (哲学社会科学版), 2010 (6): 124-130.
- [28] 石冠彬. 人工智能民事主体资格论: 不同路径的价值抉择 [J]. 西南民族大学学报 (人文社科版), 2019 (12): 94-102.
- [29] 申明详. 人性魔方 (续集) [M]. 北京: 九州出版社, 2017.
- [30] 冯洁. 人工智能对司法裁判理论的挑战: 回应及其限

- 度 [J]. 华东政法大学学报, 2018 (2): 21-31.
- (收稿日期: 2024-05-10)

作者简介:

陈思敏 (1998-), 女, 硕士研究生, 主要研究方向: 民商法学。

蔡唱 (1972-), 女, 教授, 博士生导师, 主要研究方向: 民商法学。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com