

基于数据湖平台的工业大数据分析实践： 以智能油田能效分析为例

李 满¹, 安创锋¹, 高 静¹, 牛永胜¹, 姚嘉琨²

(1. 中海石油(中国)有限公司天津分公司, 天津 300450; 2. 中国电子系统技术有限公司, 北京 100089)

摘要: 工业系统以及工业企业应用场景日益复杂, 导致系统处理数据量与数据类型日益增多。面对多样化的业务应用场景和海量多源异构数据, 对数据分析的流动性与灵活性要求越来越高。而传统基于数据库的大数据分析平台无法满足不同结构数据汇入与数据源变化。因此, 构建了一套端到端、高效协同的大数据分析应用实施框架, 结合数据湖平台与智能算法建立针对工业大数据的分析模型, 实现以业务分析需求为驱动, 结合数据湖平台对于海量多源异构数据的处理、汇聚、管理能力, 高效开展数据建模、准备、测试、训练与验证。最后, 在智能油田能效分析场景进行应用验证, 成功实现对于智能油田能效分析场景下的系统预测、优化与决策功能, 为油田全业务流程提供数据支撑, 推动智能油田可持续发展。

关键词: 大数据分析; 数据湖; 工程化架构; 智能油田; 商业智能; 人工智能

中图分类号: TP393

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2024.11.013

引用格式: 李满, 安创锋, 高静, 等. 基于数据湖平台的工业大数据分析实践: 以智能油田能效分析为例 [J]. 网络安全与数据治理, 2024, 43(11): 75-84.

Industrial big data analytics practice based on data lake platform: an example of intelligent oilfield energy efficiency analysis

Li Man¹, An Chuangfeng¹, Gao Jing¹, Niu Yongsheng¹, Yao Jiakun²

(1. CNOOC (China) Tianjin Branch, Tianjin 300450, China;

2. China Electronics System Technology Co., Beijing 100089, China)

Abstract: As industrial systems and the application scenarios of industrial enterprises become increasingly complex, the quantity and variety of data processed by the system also increase. In light of the growing number of diverse business application scenarios and the increasing volume of heterogeneous data from a multitude of sources, the need for enhanced mobility and flexibility in data analysis is becoming increasingly apparent. The conventional database-centric approach to big data analysis is inadequate for accommodating the heterogeneous structural characteristics of data sinks and the evolving nature of data sources. Accordingly, this paper presents a comprehensive, integrated and collaborative big data analysis application implementation framework. This framework combines the data lake platform and intelligent algorithms to establish an analysis model for industrial big data. Furthermore, driven by business analysis requirements, the data lake platform's processing, aggregation, and management capabilities are leveraged to efficiently carry out data modeling, preparation, testing, training, and validation. Subsequently, the application is verified in an intelligent oilfield energy efficiency analysis scenario. This successfully demonstrates the system's ability to predict, optimize, and make decisions in this context, providing data support for the entire business process of oil fields and promoting the sustainable development of intelligent oilfields.

Key words: big data analysis; data lake platform; industrial architecture; intelligent oil field; business intelligence (BI); artificial intelligence (AI)

0 引言

随着全球能源行业向数字化转型的深入发展,“智能油田”等规划相继出台,旨在通过技术创新与数据驱动,实现油气上游业务的全面升级。为此,中国海油同步立项,开展智能油田建设和勘探开发数据湖平台的建设。智能油田建设从各油田分散的业务应用场景出发,由各油田分公司根据自身特定的业务需求开展项目,以促进多样化的创新尝试和技术的快速迭代;数据湖平台作为支撑智能油田应用的数字基础设施,明确了集中化和统一化的建设方向,旨在整合中国海油上游勘探开发的核心业务数据,实现数据管理的统一化。智能油田与数据湖平台项目并行推进,随着项目的深入,两者的融合趋势日益明显,智能油田从创新探索迈向规范化和标准化,数据湖平台则在支撑应用中不断优化功能并扩展。以秦皇岛 32-6 项目为试点,成功实现了实时数据入湖与服务迁移,标准化改造数据开发过程,支持快速应用开发,为中国海油智能油田与数据湖平台的进一步融合提供了有力的支撑。

工业大数据平台作为推动制造业智能化转型的关键力量,广泛应用于生产监控、故障预测和效率优化等环节。借助物联网、云计算和人工智能等前沿技术,现有平台在数据采集、存储、分析和应用方面不断增强,为企业决策提供了坚实的数据支撑。然而,随着技术进步和市场需求增长,工业大数据平台面临数据安全与隐私保护的挑战,同时多源异构数据的增加导致传统数据库平台存在数据流通不畅和灵活性不足的问题。数据湖平台通过支撑业务场景分析,能够降低大数据分析应用开发和实施的难度,这是与智能油田项目共同追求的目标。尽管已取得一定进展,但仍需深入研究,形成可复用的标准化、工程化架构,并依托数据湖平台工具进行产品化支撑。

当前,智能油田应用面临的主要问题包括数据分散、标准不统一和业务流程复杂等。这些问题不仅增加了数据处理和分析的难度,还限制了智能油田建设的整体效率和效益。数据湖作为一种统一的数据存储池,能够容纳各种规模的结构化、半结构化和非结构化数据,提供了潜在的解决方案。目前,数据湖平台主要应用于数据管理领域,涵盖数据资产目录管理、数据源及数据处理任务、数据生命周期管理、数据治理和权限管理等功能。尽管数据湖技术在架构灵活性和开放性方面表现出显著优势,但其在性能效率、安全控制和数据治理方面仍有改进空间。将数据湖平台与工业大数据分析相结合,可以有效应对企业级的生产分析需求,并提升数据管理能力。为此,本文通过调研和借鉴大数据分析领域的相关研究成果,结合数据湖平台与智能油田应用研发的实践

经验,提出了一套基于数据湖平台的大数据分析应用实施框架。以智能油田能效分析为例,对所提出的大数据分析应用实施框架进行了应用验证。

1 大数据分析应用框架与智能油田需求分析

1.1 大数据分析应用实施参考架构介绍

大数据分析应用实施架构是一种标准化、工程化的操作流程,其核心在于依托产品化工具以及规范的操作步骤,降低大数据分析应用的开发难度,提高实施效率,并显著降低相关成本^[1]。现有的大数据分析参考架构模型主要为 CRISP-DM^[2] (Cross-Industry Standard Process for Data Mining)、SEMMA^[3] (Sample Explore Modify Model Assess) 和 KDD^[4] (Knowledge Discovery in Databases)。CRISP-DM 模型由欧洲委员会资助的联盟(涵盖 SPSS、Teradata、NCR 等公司)于 1996 年开发,代表了一种行业无关的数据挖掘架构,在各种行业的数据挖掘和大数据分析项目中得到广泛应用。CRISP-DM 是一个包括业务理解、数据理解、数据准备、建模、评估和部署六个阶段的数据挖掘过程模型。这些阶段形成了一个迭代和灵活的过程,涵盖从明确项目目标到将数据挖掘成果应用于业务决策的全过程。CRISP-DM 的优点是清晰全面,适用于多种数据挖掘项目,但可能在快速变化的大数据环境中显得缓慢。SEMMA 由 SAS Institute 开发,分为五个阶段:样本选择、探索性分析、数据修改与处理、模型构建和模型效果评估。这些阶段按线性和连续的方式组织,提供了一个清晰且结构化的分析框架,特别适合初学者和需要遵循严格分析流程的项目。但其线性流程缺乏灵活性,不适合快速迭代和适应的项目。KDD 是一个广泛认可的数据科学和数据挖掘架构,包含数据清洗和预处理、数据集成和转换、数据挖掘、模式评估和知识表示等阶段。KDD 全面且系统,提供从原始数据到有用知识转换的详细指导,但在处理大数据时可能耗时且复杂,对数据清洗和预处理有较高要求。

总结而言,CRISP-DM、SEMMA 和 KDD 是数据挖掘与大数据分析的三种主要架构。CRISP-DM 以其行业无关性和全面性著称,适用于多种项目;SEMMA 的结构化和线性流程使其易于理解和实施;KDD 则提供了一种全面系统的方法论,从数据预处理到知识发现。尽管三者对开展数据分析工作提供了方法论指导,但在实际应用中存在局限性,主要在于都过分关注以数据科学家为中心自主推进数据分析工作,而忽视了整个过程中业务人员提供精确需求定义与 IT 层面提供高质量数据供给的重要作用。因此,面对多样化的业务应用场景和海量多源异构数据,现有大数据分析应用实施参考架构模型无法满足不同结构数据汇入与数据源变化,进而难以实现高效

且灵活的数据处理与数据分析，因此需要一个端到端、高效协同的大数据分析应用实施框架。

1.2 智能油田应用实施现状与面临的问题

智能油田建设的主要目标是通过运用云计算、物联网、统计分析、深度学习等技术来提高油气勘探的准确率和生产效率^[5]。秦皇岛 32-6 智能油田项目是中国海油推进的一项创新工程^[6]，主要针对的是海上油田的数据分析应用。该项目融合了现代信息技术如大数据和人工智能，对油田的生产数据进行深入分析，这些分析包括油田生产全流程的实时监控、远程智能巡检和跨专业、跨部门的协同指挥决策。通过这些技术的应用，秦皇岛 32-6 油田实现了生产效率的显著提升和成本的有效降低，同时还推动了管理优化和组织变革。国际上，胡莱斯 (Khurais) 油田的智能化项目同样将数据分析和数字技术应用于现代油田管理。该项目利用数万个传感器收集了生产和电力消耗等方面的数据，传输到油田的“数字孪生”系统中^[7]，通过实时虚拟模型能监测、预测和诊断操作过程中可能出现的问题。工程师利用这个系统在虚拟环境中测试运营调整措施，选出最有效的方案进行部署，而不干扰实际生产操作。这项技术不仅提高了油田产量，还实现了功耗降低 18%、维护成本降低 30% 和检查时间缩短 40% 的效果，体现了大数据和智能技术在提升油田效率和减少运营成本方面的巨大潜力。

尽管国内外智能油田项目均展现出技术融合的先进性，但普遍缺乏一个统一的、跨技术领域的实施架构，能够改变各个场景和项目导向的工作模式^[8]。此外，大数据和人工智能技术的快速发展与迭代更新，也使得形成一个稳定且通用的实施架构变得更加困难^[9-10]。随着多源异构数据数量与类型不断增加，同时多样化的业务应用场景需要灵活的数据分析，而传统数据库难以应对海量多源异构数据和系统分析与决策的实时性要求，亟需一个适用性广泛的端到端、高效协同的大数据分析应用实施框架，以满足现代工业系统灵活、高效的数据处理与系统决策需求。

2 基于数据湖平台的工业大数据分析应用实施架构

为了实现从数据采集、处理到分析应用的全流程，本文构建了一套端到端、高效协同的大数据分析应用实施流程，如图 1 所示，紧密依托数据湖平台，实现精准落地与广泛应用。该流程以业务分析需求为驱动，标准化数据的采集、治理、汇聚和服务，确保高质量数据供给。利用平台工具集，高效开展数据建模、准备、测试、训练与验证，将业务需求转化为可操作的数据分析应用，实时集成于生产环境，提升决策支持能力，为工

业系统的全业务流程提供数据支撑，推动工业大数据分析平台的发展。

2.1 架构总览

工业大数据标准应用实施架构模型能够有效适应业务分析需求的快速变化和大数据分析技术的发展，推进数据湖平台智能化决策建设，并实现模型信息的获取与共享^[11]。总体架构框架包括业务模块、分析模块和数据模块。业务模块从业务分析需求出发，以数据分析应用 APP 的开发和实施为核心，结合实际生产数据进行应用部署并优化分析模型；分析模块将业务需求转化为商业智能 (Business Intelligence, BI) 和人工智能 (Artificial Intelligence, AI) 分析模型，通过数据湖技术供给实际生产数据，并进行模型训练、测试和验证；数据模块基于元数据管理和数据标准化管理技术构建数据模型，并与相关数据库表结构和分析模型建立关联，同时根据分析模型训练数据的需求，进行数据采集、治理、汇聚和服务的全链路管理与数据资源管理，从而实现高效的数据管理和分析，促进整体系统的应用和优化。

基于数据湖平台的大数据分析应用端到端实施架构的核心步骤包括：(1) 精准定义业务需求，来确保数据分析与油气田企业生产全业务流程的紧密对接；(2) 构建并迭代优化数据分析模型，使之能够有效地将业务需求转化为可操作的数据分析建模任务；(3) 实施统一数据标准管理和数据供给策略，利用数据湖平台数据治理与服务工具确保供给侧数据的有效管理与共享。

2.2 需求定义

对应图 1 所示 BI. 业务分析需求定义，需要有一套模板对需求进行精确描述，以降低沟通成本^[12]。参考 CRISP-DM 并结合项目实践，给出的业务分需求定义架构模板依次包括分析主题/分析对象、价值存在/分析目的、所需“输入”、期望“输出”、前提假设/可行性评估五个模块，如图 2 所示^[13-14]。首先，定义分析对象或分析主题，针对所定义的分析对象确立分析目的；其次，建立分析对象的分析指标及统计维度，根据分析指标与知识支撑确立数据采集范围；再次，根据分析指标及统计维度确立描述型式，给出问题原因分析，并提出相应解决方案建议与预测型分析指标；最终，给出系统可行性评估及相应指标。

需求定义应严格遵循“可追溯、可量化、可实现”的“三可”原则，确保分析主题明确、价值清晰、输入具体、输出量化，并全面评估业务逻辑与数据质量，以支撑高效、精准且可行的数据分析项目。通过数据采集手段获取数据，并借助知识支撑为分析提供理论基础。在数据处理阶段，强调了对数据维度、业务逻辑、范围

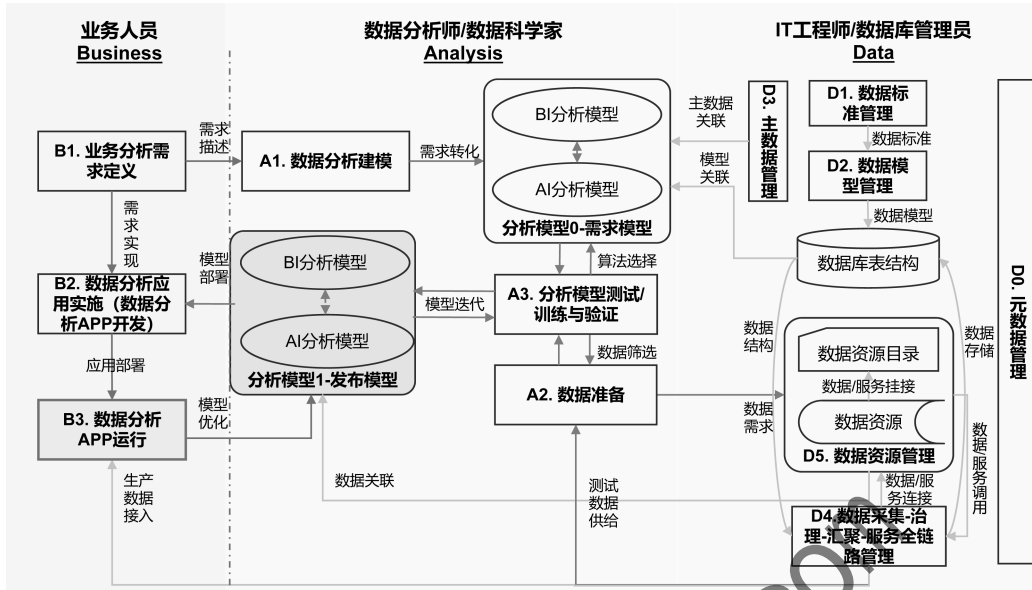


图1 基于数据湖平台的大数据分析应用端到端实施架构流程

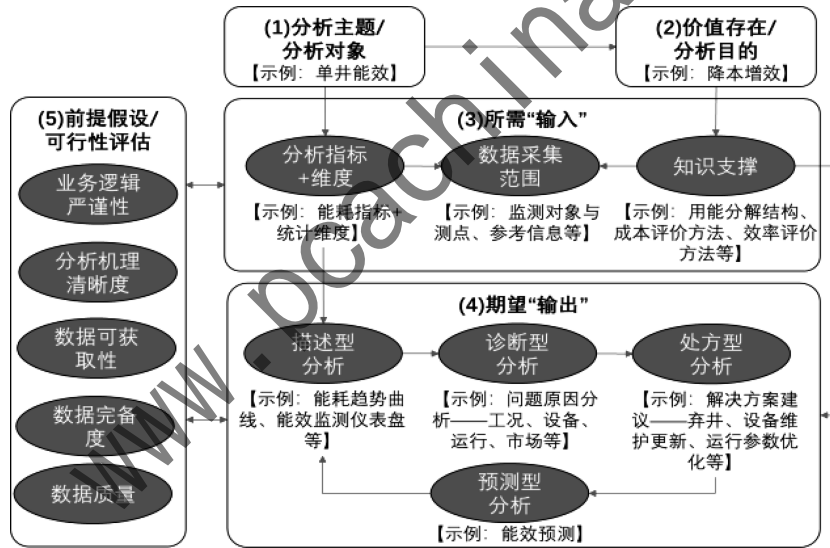


图2 需求定义标准化架构

和严谨性的把控，以确保数据的准确性和分析的深度。利用多种分析方法，对数据进行全面剖析。在整个分析过程中，注重数据完备性和数据质量的监控，以确保分析结果的可靠性和有效性。

2.3 数据采集

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。因此，数据采集侧的采集、治理、汇聚、服务全链路数据加工处理流程以数据资源

的标准化治理与服务为目标，包括：数据计算、数据全量入湖、数据规范、数据聚合、应用支撑等；另外，服务过程采用 Flink 实现，包括：数据读取、数据计算、数据推送、实时服务推送等。在数据供给侧，IT 工程师/数据库管理员依托数据治理与服务工具维护数据分析所需的各种数据资源。

这一过程通过一系列关键环节实现：元数据管理确保数据元信息的清晰可追溯；数据标准管理奠定数据一致性与准确性的基石；数据模型管理优化数据架构以支持复杂分析；主数据管理保障关键业务数据的权威性与准确性；数据采集至服务全链路管理构建数据从源头至

应用的无缝流转体系；数据资源管理强化数据资产的全面监控与价值最大化，构成数据从源头进入数据湖的标准化采、存、管、服自动加工处理工厂，持续为数据分析应用提供高质量的数据资源。所提出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、数据管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。具体数据供给侧数据加工处理流程展示在图3中，并且所提出的全链路数据加工处理流程已在数据湖平台二期项目付诸实现。

2.4 分析建模

数据分析建模是基于业务和数据的理解，选择合适的算法和建模工具，对数据中的规律进行固化、提取，最后输出数据分析模型。数据分析建模包括 BI 分

析与 AI 分析，两者相辅相成。BI 和 AI 的分析建模的架构流程，在数据分析的需求侧与供给侧之间，数据分析师及数据科学家首先解析业务部门提出的分析需求。利用数据湖平台提供的标准化模型、主数据以及业务数据，结合该平台集成的高级分析工具，包括但不限于业务智能 (BI) 和人工智能 (AI) 工具。通过采纳“模型加数据”策略，开展了数据建模、数据准备以及分析模型的测试、训练和验证，确保所开发模型的准确性和可靠性。此过程也确保了业务场景的具体需求通过数据分析得以精准满足，从而实现了数据分析的有效供需对接。

2.4.1 BI 分析

BI 分析常常采用描述型分析手段来完成模型搭建，而基于数据仓库开展可视化图表建模工作，通常列属于基础分析应用。架构流程的实践步骤如图4所示。

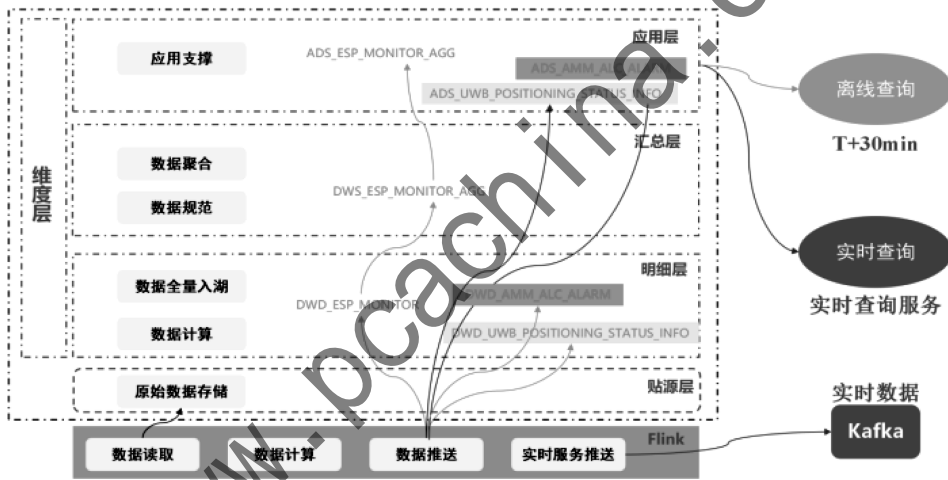


图3 数据供给侧全链路数据加工处理流程



图4 BI分析建模、测试验证与应用部署架构流程

(1) 指标建模: 基于多维数据仓库构建指标 + 维度星型 - 雪花型模型。

(2) 指标计算与预警/报警阈值设定: 针对每个指标定义计算方法、统计口径、时间频次、数据来源、阈值定义。

(3) 相关参数设定: 设定/接入用于指标计算与对标分析的参数值, 如实时电价、国际油价等。

(4) 可视化分析图表设计: 设计用于分析展现的图表模板与组件, 包括趋势图、散点图、饼图、直方图、仪表盘等。

(5) BI 分析测试验证: 使用自助分析工具在测试环境基于可视化分析图表接入测试数据, 定制 BI 分析结果综合展示界面, 如管理驾驶舱、生产监控大屏等, 可视化展现趋势分析、对比分析、帕累托分析等统计分析曲线、图表, 并能进行上卷、下钻操作。

(6) BI 分析应用部署: 将指标模型与 BI 分析结果综

合展示界面发布部署到生产环境, 接入真实生产数据。

2.4.2 AI 分析

AI 分析一般对应诊断型、处方型和预测型分析, 基于湖仓一体的大数据资源开展数据融合与知识发现建模工作, 属于高级分析应用。相比 BI 分析, AI 分析要复杂得多, 可参照 IDEF0 来描述 AI 分析的基本原理, 如图 5 所示, 将问题求解过程转化为输入 (I)、处理 (P)、输出 (O), 其中输入为原始数据, 输出即为经问题求解处理后生成的知识/洞察, 为实现这个过程, 需要有控制 (C) 和机制 (M), 其中控制对应算法, 机制则对应模型。AI 分析需要数据、算法、模型、知识, AI 分析过程即数据库、算法库、模型库、知识库这“四库”关联迭代的过程。

AI 分析建模、训练验证与模型部署的架构流程如图 6 所示, 分步描述如下。

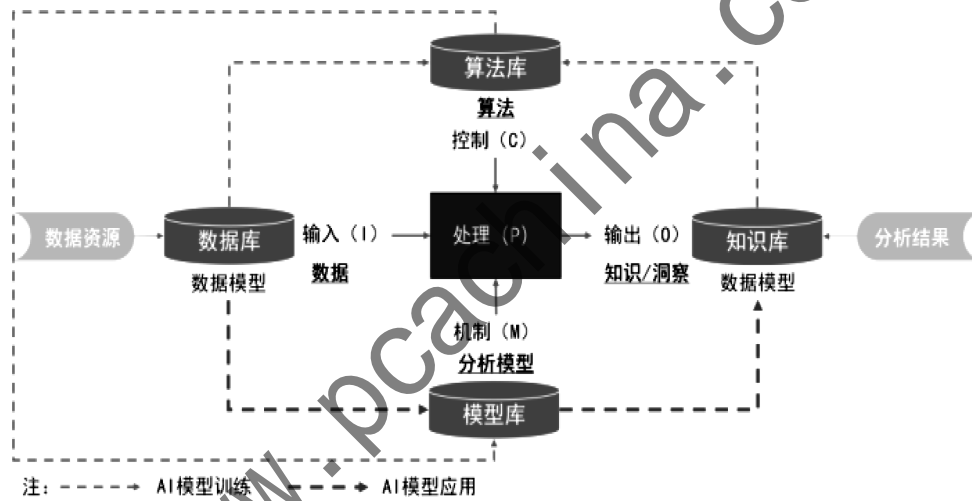


图 5 AI 分析过程“四库”联动机理——IDEFO 模型描述



图 6 AI 分析建模、训练验证与模型部署架构流程

(1) 需求建模：根据不同 AI 分析场景建立 AI 分析的问题求解模型，如分类问题、聚类/模式识别问题、关联问题、推理问题、预测问题等，并针对需求模型的输入 (I) 和输出 (O) 关联相应的数据模型与主数据，此时建立的模型是针对业务与数据理解产生的面向问题定义的 AI 分析问题描述模型，尚未关联具体算法 (C) 来建立求解模型 (M)。

(2) 数据准备：根据需求模型从数据资源目录中选择并定制数据集，包括原始数据与分析结果数据，并接入测试环境以获取测试数据，将测试数据分为两份，一份作为训练数据，用于训练分析模型；另一份作为验证数据，用于对训练好的分析模型进行评价验证。

(3) 模型训练与验证：从算法库中选择合适的算法，并对定制数据集中的测试数据进行主成分分析和特征提取，完成对数据进行筛选，然后使用训练数据跑算法进行分析模型的训练，以基于不同的算法训练实现不同模型的生成，然后使用验证数据对不同模型进行评价验证，在此过程中需要进行算法、数据和模型的多轮迭代，最终选择效果最好的模型提交发布。

(4) 模型部署：发布的模型部署到生产环境，接入生产环境的真实数据，对业务用户提供数据分析应用服务，并在使用过程中不断积累历史数据，基于历史数据对 AI 模型进行训练，以实现 AI 模型的持续优化。

3 实例验证：智能油田能效分析

为了验证所提出大数据分析应用实施框架的有效性，本文以智能油田能效分析为例，验证所提出基于数据湖平台的工业大数据分析架构的适用性，并详述数据湖驱动的智能决策方法。通过依托数据湖平台，能够实现从监测到数据整合、模型建立的端到端流程，为能效分析提供坚实技术支持。所提出的架构规范操作，融合大

数据分析，显著提升油田运营效率。

3.1 实例总述

智能油田是深度融合物联网、大数据、云计算与 AI 技术，实现全面感知、自动操控、精准预测与优化决策，促进油田高效、绿色、可持续运营的数字化与智能化综合系统。在智能油田能效分析需求侧的实施过程中，首先从能效数据供给入手，通过智能油田数据服务进行标准化，并存储至能效分析数据库中，确保数据的准确性、完整性和实时性。随后基于系统数据建立能效分析模型，对油田的耗电量等能效数据进行线性回归分析，建立了 BI 分析模型，得到了能效问题的分析结果和相关建议。最后，对 AI 分析部分进行了深度学习工具的引入，进一步完善了应用架构的结构。通过使用 AI 模型部署能效数据分析 APP，实现了对问题的实时分析与诊断，为提供优化决策方案奠定了基础。油田能效分析流程如图 7 所示。

3.2 需求定义

为提升智能油田的单井能效，实现对能源的精细化管理这一作用价值，本文根据提供的架构对智能油田能效分析解决方案进行业务需求分析，分析方法如图 8 所示。首先，由业务人员确定油田能效分析的分析指标，从宏观综合指标、工艺过程指标和终端设备指标三方面进行能效指标梳理，确定能效数据的指标及统计维度。其次，确定能效数据对应的采集范围，例如油田设备的耗电量、产能值及日产液量等。然后，设计智能油田服务标准化方案，对能效数据进行处理，同时确定耗电成本分析方法及能效评价标准，用于后续的能效数据分析。最后，提出耗电量 - 产能可视化结果分析需求，以及对油田工作异常问题的定位。根据异常诊断结果获取油田单井进行优化建议，得到油田的能效水平发展趋势进行预测与决策方案。

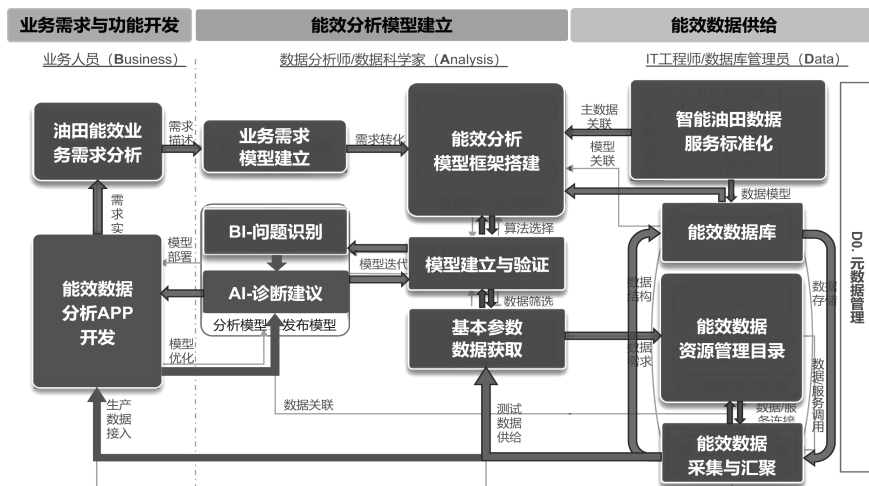


图 7 油田能效分析流程

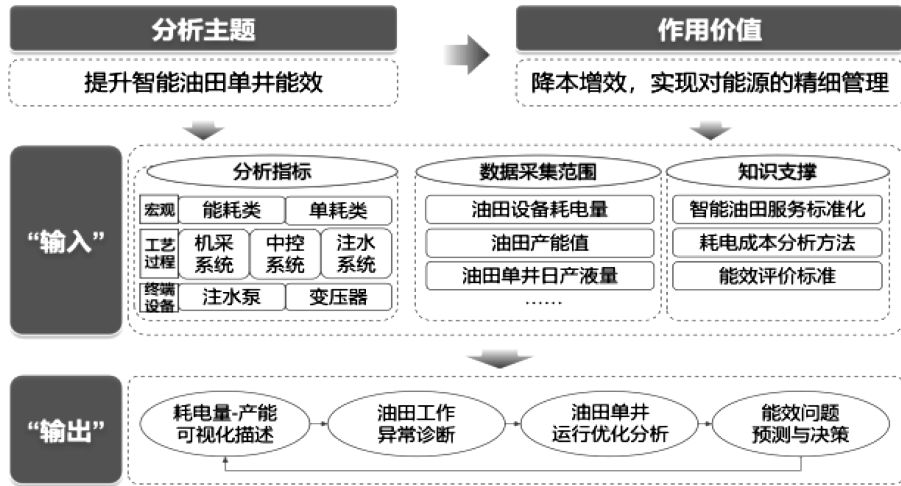


图8 油田能效业务需求分析

3.3 数据采集

在数据供给方面, 通过从物联网平台上的各种油田数据源中采集能耗值、产量值等能效相关数据, 设置资源管理目录, 如表1所示。根据智能油田数据处理要求, 构建Kafka集群、Flink计算引擎以及CeaSQL-DW实时数仓, 建立数据服务标准化模型。数据主体部分基于数据湖推送至Kafka集群, 由Flink从Kafka集群中提取实时数据进行处理, 同时关联客户业务维度数据库和历史数据库, 依据数据服务标准化架构进行实时数据建模, 对实时监控数据进行聚合计算, 同时开发和发布服务工具。最后, 利用数据湖平台工具实现实时数据服务的全链路监控管理, 以确保整个数据流程的稳定性和可靠性。

表1 能效分析指标管理示例

分类	序号	数据集名称	数据集标识
生产参数	1	注水泵监测	DWD_DCS_WATER_INJ_
	2		PUMP_MONITOR
	3	控制器监测	ADS_SWI_MONITOR
	4		ADS_ESP_MONITOR
	5	中控系统	ADS_ESP_MONITOR
能耗管理	1	时度汇聚表	Ads_Ems_Agg_Hourly
	2	日度汇聚表	Ads_Ems_Agg_Daily
	3	月度汇聚表	Ads_Ems_Agg_Monthly
	3	年度汇聚表	Ads_Ems_Agg_Yearly

3.4 分析应用

在油田能效数据分析中, 油田能效数据库包含不同设备的生产信息、油井的能源消耗等, 关联信息为后续

的能效预测工作提供了可靠的数据基础。在实施基于油田能效分析的大数据架构时, 首先结合从油田能效数据库中获取的耗电量和产能值等历史数据, 通过数据清洗和预处理, 确保数据的质量, 接着运用线性回归的数据分析方法构建BI预测模型, 得到了耗电量与产量之间的函数关系。然后设置一系列基本参数, 如美元汇率和原油密度等, 用于模型测试并形成可视化界面。利用实时的能效数据对模型进行测试验证。

在油田能效分析中, 实时准确地预测油田耗电量对于识别能源使用中的低效环节以及优化电力资源分配至关重要。本文提出了一种基于生产数据的耗电量预测模型, 实现了对油田耗电量的实时预测。该模型能够自动提取油田能效数据库中的生产参数作为输入, 并通过模型计算输出油田的预估耗电量。为验证预测结果的准确性, 对比分析模型预测的耗电量与实际耗电量, 并计算各油井实际耗电量与预估耗电量的均方根误差 (Root Mean Squared Error, RMSE), 如图9所示。结果显示, 油井的预估耗电量与实际耗电量十分相符, 均方误差为10.802, 能够精确预测油田的实时耗电情况。

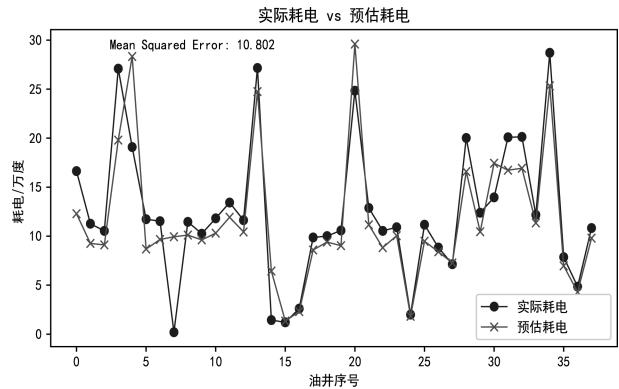


图9 预测耗电量与实际耗电量

为了进一步提升油田能效管理水平,本文引入了基于专家系统的 AI 技术进行系统异常分析和预警。对于所有生产数据均处于预警区的油井或是运行突然中止的油井, AI 技术将其确定为异常油井并生成报警;对于多项生产数据处于预警区的油井,将其确定为可能异常的油井,并生成警告。例如 16 号和 25 号油井所有生产数据均处于预警区,如图 10 所示, AI 系统将其认定为异常油井,生成报警,并提供针对性的检查建议。16 号油井虽然正常运行,但耗电量较低;25 号油井则在工作 311.5 小时后停止了运行。针对这些异常情况, AI 系统

建议进行实地勘测,以确认设备是否正常,并查明潜在问题。对于多项生产数据处于预警区的油井的, AI 系统将其认定为可能异常的油井,生成警告,如 15 号、28 号、36 号以及 37 号油井。针对这些可能异常的情况, AI 系统将重点监控其处于预警区的生产数据,多项生产数据一直处于预警区,将升级为异常油井。根据 AI 模型的预测结果,系统还可以生成用能分析建议,如对单井进行经济性评估、停用或废弃等。这些建议有助于优化油田生产设备,提升油井的能效水平,从而改善整体能效管理。

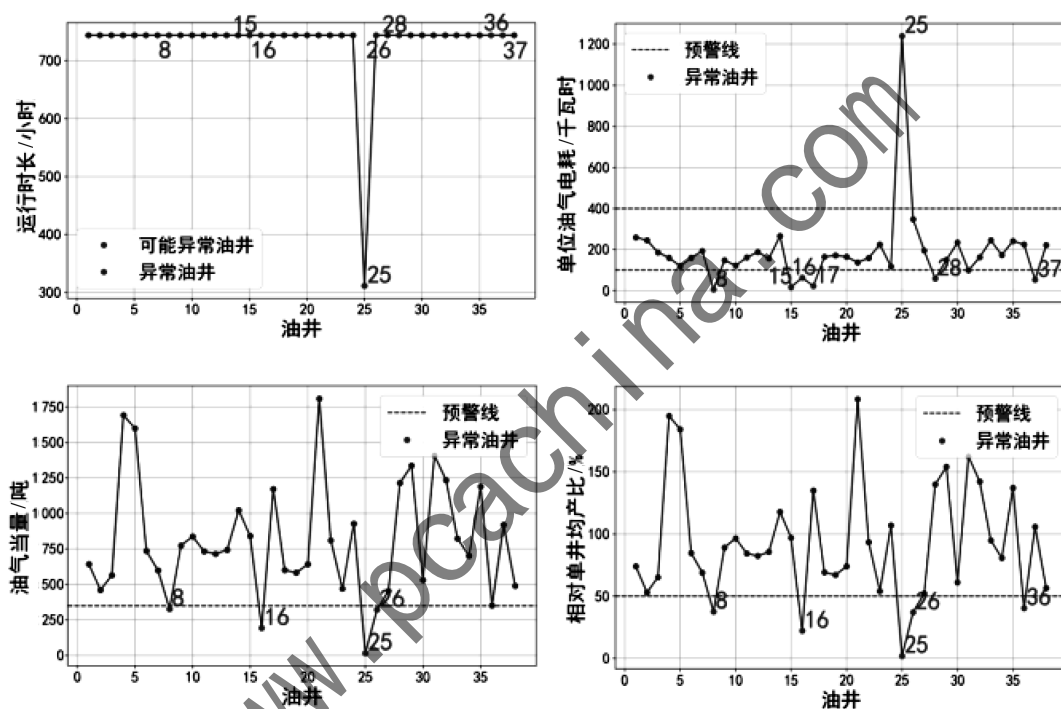


图 10 油井的生产数据

4 结论

本文提出了一套面向智能油田的基于数据湖平台的大数据分析应用实施架构。该架构通过系统化的数据供给、精准分析建模与应用开发三大核心阶段,形成了完整的端到端解决方案。使用所提出的基于数据湖平台的大数据分析应用架构,不仅降低大数据分析的实施难度,还显著提升了工业大数据平台的运营效率和管理水平。验证结果表明,采用这一架构能够实现数据资源的高效管理和分析模型的持续优化,从而为智能油田的智能化决策提供强有力的支持。同时,所提出的大数据分析应用实施架构还适用于智能油田的其他应用场景,可以实现油气田更加精细化的管理和决策,推动油气田管理从经验驱动向数据驱动的转变,为油气田的智能化管理提供了重要的借鉴意义。

参考文献

- [1] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和数据分析技术综述 [J]. 软件学报, 2014, 25 (9): 1889-1908.
- [2] 黄海. 基于 CRISP-DM 和 PCA 数据分析技术在制造业企业的一个应用实践 [J]. 工业控制计算机, 2021, 34 (10): 138-140.
- [3] 钟颖, 王松, 吴浩, 等. 基于 SEMMA 的网络安全事件可视探索 [J]. 浙江大学学报(理学版), 2022, 49 (2): 131-140.
- [4] 陈福集, 杨善林. 一种基于 KDD 的 Web 搜索引擎框架 [J]. 情报学报, 2002, 21 (3): 264-268.
- [5] 陈绍凯, 安鹏, 张能, 等. 领域驱动设计在中国海油智能油田建设中的探索与实践 [J]. 中国海上油气, 2023, 35 (6): 189-196.
- [6] 于杰, 王保平, 郭家全. 克拉玛依油田: 中国石油两化融合的信息化产业硅谷 [J]. 中国管理信息化, 2014, 17

- (13): 47-52.
- [7] 王子宗, 高立兵, 索寒生. 未来石化智能工厂顶层设计: 现状, 对比及展望 [J]. 化工进展, 2022, 41 (7): 3387-3401.
- [8] 郭福军. 油田企业经济评价数字化模式探讨 [J]. 石油科技论坛, 2023, 42 (3): 72-78.
- [9] 李阳, 廉培庆, 薛兆杰, 等. 大数据及人工智能在油气田开发中的应用现状及展望 [J]. 中国石油大学学报 (自然科学版), 2020, 44 (4): 1-11.
- [10] 夏博强. 大数据、物联网及人工智能技术在海上智能油田生产中的应用研究 [J]. 数字通信世界, 2023 (2): 121-123.
- [11] 何文韬, 邵诚. 工业大数据分析技术的发展及其面临的挑战 [J]. 信息与控制, 2018, 47 (4): 398-410.
- [12] 叶康林. 基于实时数据库的数据可视化分析系统 [J]. 计算机仿真, 2017, 34 (1): 129-131.
- [13] 工业互联网产业联盟大数据系统软件国家工程实验室. 工业大数据分析指南 [M]. 北京: 中国工信出版社, 电子工业出版社, 2019.
- [14] 贾雪琴, 罗松, 胡云. 工业互联网标识及其应用研究 [J]. 信息通信技术与政策, 2019, 45 (4): 1-5.
- [15] 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.
- (收稿日期: 2024-09-14)

作者简介:

李满 (1984-), 男, 本科, 工程师, 主要研究方向: 油气行业信息化、智能化建设。

安创锋 (1978-), 男, 硕士, 高级工程师, 主要研究方向: 海上油气生产管理、油气行业信息化、智能化管理。

高静 (1979-), 女, 本科, 高级工程师, 主要研究方向: 油气行业信息化、智能化管理。



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