

数据管理对采油厂 数据建设的意义探讨

文◆大庆油田有限责任公司第四采油厂数字化运维中心 杜 屹

引言

随着油田企业改革和数字化技术的应用和深入，数据集成和管控变得愈发重要。面对接踵而至的海量数据，有效、合理的数据运维工作，不仅能够降低生产和管理成本，使管理者及时做出生产经营决策，还能够提高企业经济效益，增加企业核心竞争力。数据作为油田生产运营的核心要素，在生产决策过程中发挥着至关重要的作用，不仅能够为决策者提供全面的生产信息，还能够为优化生产流程、提高生产效率提供有力支持。

1 数据发展现状

信息技术在采油厂的发展主要体现在数据采集、数据存储和管理、数据处理、数据分析和解释、多学科协同工作、信息共享和交流、决策支持等方面。目前，随着采油厂数字油田建设的不断推进，各类软件系统研发和数字化设备种类逐渐增加，数据呈现出结构复杂化、来源多样化、增长高速化的特点。为满足

油田未来海量数据传输、计算和存储的实时性需求，应建立一套完整的数据管理机制，实现缩短业务人员收集、整理有效数据的时间，进而缩短研究周期，进一步提升管理效率。

2 数据管理方式研究

2.1 数据运维方向研究

为了适应发展需求，探索一套行之有效的数据运维方法是数据治理及应用的关键，数据运维主要体现在以下 5 个方面。

(1) 提升数据质量。数据质量是数据的综合表现。在采油厂数据库中，数据主要来源于生产数据的上报和回迁、系统产生的日志数据、管理类软件产生的数据、数字化设备产生的实时数据等。为了使数据中的各种问题对使用情况影响最小化，应针对数据发生、使用的各个节点进行管控，从而达到对生产情况有的放矢进行数据支持的目的。

首先是数据预处理。随着数字化项目地推进，采集的实时数据量在每日的数据交互中占比可观，未来该数据量将会呈线性增长。由于受到网络、硬件设备、实际生产情况、采集设备的影响，其中存在大量冗余数据，而冗余数据会对后续的数据利用形成操作障碍，数据清洗将会大大削减脏数据对于生产情况分析的影响，从而高效发挥数据自身的价值^[1]。其次是对数据的真实性和完整性分析。数据资源作为一种隐形资产，在油田企业中有着不可或缺的重要地位。数据要“真”，也要“全”，数据的真实性和完整性是数据应用的基础。对重要数据库进行数据逻辑验证、数据格式校验、数据缺失判断、完整性检测以及从传输损坏、硬件故障、配置问题、人为错误、故意破坏等方面对数据进行监控是数据运维工作的重要组成部分。

(2) 数据库性能监控。数据库性能的优劣将直接影响数据的使用效率，通过扩大数据库存储量（即数据总量和 IO 吞吐量）、扩大数据库会话数量、提高 CPU 利用率以及优化内存使用情况等方面，实现对数据

【作者简介】杜屹（1989—），女，黑龙江北安人，本科，中级工程师，研究方向：软件开发、数据管理。

库性能监控。同时，应监控采油厂各个程序运行效率，优化慢查询语句，提升数据库性能。

(3) 数据安全监控。以数据为目标的网络攻击已经成为数据时代新的安全威胁。采油厂信息安全需求已经从基本的网络安全防护升级为对数据的安全防护。因此，针对数据运维安全，应了解数据资产的分布情况，实现对数据风险的及时预警。

(4) 数据流向监控。对数据进行全生命周期管理、溯源到数据发生的各个节点，有助于采油厂建立数据和人工智能算法互联互通的生态^[2]。通过监控数据流向，实现数据来源跟踪，最终实现数据影响分析以及任务依赖分析。

(5) 硬件设备数据监控。一是对数据库环境数据进行监控，检测温湿度等关键数据。二是对自动化设备性能进行监控，监测设备的在线率、完好率，在数据收集和存储方面达到基础性完善，提升数据存储的可靠性。

2.2 数据分级技术研究

(1) 数据分级处理的基本原则。工业数据的分级治理是确保数据安全的基准线。随着越来越多的设备系统生产和服务暴露在互联网中，数据安全风险越来越大。为此，应站在安全的视角上进行分级，结合工业数据的属性、安全防护要求、构建数据分级的制度体系和分级施策，确保工业数据的安全^[3]。数据分级应遵循以下原则，一是实时性原则。油田实时数据分级处理需要保证数据的实时性，确保数据的及时传输和处理，为油田生产管理提供实时、准确的数据支持。二是完整性原则。数据分级处理应确保数据的完整性，防止数据在传输和处理过程中丢失或损坏，保证数据的可靠性和可用性。三是安全性原则。数据分级处理应考虑数据的安全性，对不同级别的数据进行不同的安全措施，防止数据泄露和恶意攻击。四是可扩展性原则。数据分级处理应具备一定的可扩展性，能够根据油田生产的实际需求，灵活调整数据分级策略，满足不同场景的数据处理需求。

(2) 数据分级处理的具体方法。数据分级标准。一是明确工作构想和数据分级范围。二是从数据生成系统和业务出发，将业务框架、逻辑和表达过程按照业务、系统、模块等尺度进行分类。三是从业务和系统方面观察数据分布情况，明确数据在哪里、数据由谁提供、数据格式是什么等方面的问题。数据分级按照一定的原则和方法，主要目的是便于数据开放和共享，同时也是数据保护工作中的关键部分，为制定安全、准确、完善数据策略提供数据支撑。

根据油田实时数据的类型，将其分为关键数据、重要数据和普通数据3个级别。关键数据包括井口产量、井口压力等关键生产数据；重要数据包括站间流量、温度等重要监控数据；普通数据包括环境参数、设备运行状态等一般性数据。根据油田实时数据的来源，将其分为井场级数据、站间级数据和油田级数据3个级别。井场级数据包括井口产量、井口压力等数据；站间级数据包括站间流量、温度等数据；油田级数据包括油田的生产运行数据。根据油田实时数据的处理方式，将其分为实时处理数据和批处理数据两个级别。实时处理数据是指对实时采集

到的数据进行实时分析、处理和应用；批处理数据是指对一段时间内采集到的数据进行集中处理、分析和应用。根据油田实时数据的安全要求，将其分为公开数据、内部数据和机密数据3个级别。公开数据是指不涉及油田生产安全的公开数据；内部数据是指油田生产管理所需的内部数据；机密数据是指涉及油田生产安全、商业机密的核心数据。根据数据的时效性，将实时数据分为实时数据、近实时数据、历史数据等^[4]。

(3) 数据分级技术。目前，随着实时数据量、数据维度的扩展，采用自动分级的方式，补足人工分级。自动分级中，常用的技术方法有以下3种。一是特征选择和降维。通过特征选择和降维技术，从原始数据中选择最具代表性的特征，如主成分分析、线性分析等。二是时间序列分析。在实时数据的分级中，通过时间序列分析，揭示时序数据的趋势、周期性、季节性等规律。三是自然语言识别和图像识别。针对于文本类数据、图像数据的分级具有良好的适配性。

2.3 实时数据管理研究

(1) 自动采集数据特点。对多个不同的生产场景的采集数据进行观测可知，其存在以下4种数据特征。一是影响因素的多样性。在实际生产中，影响数据质量的原因较多，采集的质量问题、传输的质量问题、存储的质量问题等，不同生产场景的数据质量不尽相同，各个方面的影响因素同时存在。二是行为发生的延迟性。数据质量欠佳的情况有时候不会突然表现出来，由历史因素的积累、平台同步或者工具

技术的缺陷等因素导致。三是多种数据的相关性。自动采集获取的数据具有很强的相关性，借由集输系统的流程节点，可获悉多种类采集数据的逻辑关系，并根据生产经验判断其合理性。四是采样时间的差异性。由于实际采集数据类别多元，采集设备不尽相同，采集频率各有差异，因此呈现出多种时间尺度，如毫秒级、秒级等。

(2) 数据质量评价现状。现有的数据评价方法大多依赖于数据提供者或使用者的主观判断。人工形式观测数据时，有明显逻辑错误的易被观测，异常值变动不明显的难以被发现，该方式存在一定的滞后性。同时，由于数据采集的密度大，在追溯低质量数据产生的时间点时，产生的工作量也较大，且没有统一的度量标准，也会影响异常情况的判别。

(3) 数据质量评估方法。建立数据质量控制体系，从多角度、多方面明确数据质量的检验标准。从采集数据的提供者、生产者、使用者等多个角度出发，对数据质量进行有效评估，提高生产数据对使用者需求的满足程度。全面的数据质量管理，需要每一位信息使用者和维护者共同参与，形成数据质量的闭环管理以及循环优化。

(4) 数据质量的约束条件。根据自动采集数据的特性，结合油田生产数据的实际情况，对采集数据的质量从9个维度进行约束，数据质量评价指标如表1所示。

(5) 数据质量的控制标准。对数据完整性、准确性、适度性、及时性共4个维度的指标进行量化计算，对时效性、适用

表1 数据质量评价指标

评价角度	定义
完整性	数据是否无缺省，在广度和深度上的充分程度
准确性	数据准确程度
适度性	对某个具体任务来说，数据量适度程度
及时性	数据从采集端到入库是及时程度
时效性	对某一生产场景来说是否是最新数据
适用性	数据贴近生产实际程度
可信性	数据来源真实可信程度
一致性	数据以同样格式表现的程度
安全性	数据存取和维护的途径安全程度

性、可信性、一致性、安全性共5个维度指标进行定性分析。

3 数据管理目标

(1) 实现数据资源的统一管理。在采油厂数据建设中，数据管理的首要任务是实现数据资源统一管理。实现这一目标主要着眼于解决数据来源驳杂、数据重复采集、数据标准不统一等现实问题。制定统一的数据管理规范和服务机制，确保各类数据按照统一的标准进行采集、存储和处理，积累有效的数据资产。

(2) 实现数据资源优化调配。在数据资源统一管理的基础上，数据管理还应实现数据资源的优化调配。旨在促进信息的纵横贯通，构建完整的数据生态，打破不同专业、部门之间的信息孤岛和业务孤岛。深入分析数据资源的分布和使用情况，根据业务需求进行数据资源的合理配置和调度，确保数据资源得到充分利用和发挥最大的价值。

(3) 实现数据性能监控。数据性能监控是数据管理的重要目标之一。实时监控和分析数据资源的性能，及时发现和解决数据运行过程中的问题，有利于确保数据的稳定性和可用性。同时，数据性能监控还能够为硬件资源的优化配置提供有力支持。分析数据处理的性能瓶颈和资源利用率，探索出合理的区域配置模式，降低硬件资源的投资成本和运维成本。

结语

数据作为新兴资产，其在采油厂数据建设中的价值日益凸显。通过有效的数据管理，不仅能够实现数据资源的统一采集、统一管理和统一服务，还能够提高采油信息共享能力和对专业应用无缝支持的能力。不仅能够充分发挥数据的职能作用，还能够为采油厂数据中心的建设提供有力保障，推动采油厂的数字化转型和智能化升级。

引用

[1] 商希雪,韩海庭.数据分类分级治理规范的体系化建构[J].电子政务,2022(10):75-87.
 [2] 孔德明.数据财产权到访问权:欧盟数据设权立法转型解析[J].比较法研究,2023(6):33-50.
 [3] 熊丙万.论数据权利的标准化[J].中外法学,2023,35(5):1145-1164.
 [4] 陈兵.科学构建数据要素交易制度[J].人民论坛·学术前沿,2023(6):66-78.