

油田生产数据治理技术及应用研究

王楠 (辽河石油勘探局有限公司信息工程分公司, 辽宁 盘锦 124010)

摘要: 针对油田生产数据治理问题, 本次研究首先对油田生产数据的特点进行分析, 对油田生产数据治理工程建设存在的问题进行探讨, 最后, 开展油田生产数据治理技术及应用研究, 为推动油田生产数据治理领域的进一步发展奠定基础。研究表明: 油田生产数据治理对于确保数据的完整性以及保障油田企业的经济效益都十分关键, 在开展油田生产数据治理工程建设的过程中, 主要存在缺乏统一标准的数据共享、数据处理技术落后、专业人才匮乏等问题, 因此, 需要根据数据治理的流程, 选择合理的数据治理方法, 把控数据治理的要求, 进而使得油田生产数据治理质量全面提升。

关键词: 油田; 生产数据; 特点分析; 数据质量; 应用研究

0 前言

随着现代科技的不断发展, 油田生产行业也正经历着革命性的转型, 这一转型不仅关乎石油和天然气资源的更有效利用, 还涉及到油田生产数据的广泛应用和管理, 油田生产数据是石油行业的宝贵资产, 它们记录了地质勘探、钻井、生产和设备状态等多个方面的关键信息, 这些数据不仅对生产决策和效率的提升至关重要, 还对环境保护具有重要意义。有效的数据治理技术可以帮助油田生产企业更好地监控生产过程, 识别潜在问题, 并采取及时措施, 有助于提高生产效率, 减少停工时间, 降低生产成本, 数据分析和机器学习技术可以帮助企业更好地理解油田生产数据, 识别潜在的优化机会, 从而更有效地配置资源, 减少浪费^[1]。

由此可见, 数据治理工作对于油田的发展十分关键, 本次研究主要是对数据治理存在的问题和数据治理流程、方法、要点进行分析, 以此使得数据治理质量全面提升。

1 油田生产数据的特点分析

油田生产数据来自多个来源, 包括地质勘探、钻井、生产、环境监测等多个领域, 这些数据可以包括数字数据、文本数据、图像和视频等多种类型的信息。部分油田生产数据需要实时监测和处理, 以及时应对生产问题和安全风险, 例如钻井和井口操作需要实时数据以确保安全和生产效率。油田生产数据通常具有大规模性, 因为油田生产过程复杂, 需要大量数据来记录各种参数和指标, 这使得数据的存储和处理成为挑战。油田生产数据来自多个油井、设备和传感器, 分布在广泛的地理区域, 数据的多源性需要采用多种技术来进行数据采集和集成。很多油田生产数据是时

序数据, 记录了随时间变化的信息, 这些数据包括生产率、压力、温度、设备状态等, 需要进行时间序列分析^[2]。一些油田生产数据需要高分辨率的数据采集, 以捕捉细微的变化, 例如地震数据和勘探数据需要高分辨率来揭示地下结构。油田生产数据通常具有多个层次, 从地下地质层次到地表设备层次, 这需要多层次的数据管理和分析, 油田生产数据的复杂性体现在数据之间的关联性和多变性, 不同数据之间可能存在复杂的关系, 需要高级的数据分析技术来理解。油田生产数据通常包含敏感信息, 如井位、勘探结果等, 因此数据安全和隐私成为关键问题, 需要采取措施保护数据的机密性^[3]。

2 油田生产数据治理工程建设存在的问题

2.1 缺乏统一标准的数据共享

油田生产数据涵盖多个领域, 包括地质、勘探、钻井、生产、环境等, 每个领域都可能有自己的数据格式和标准, 这种多样性导致了不同类型的数据难以整合和共享。油田生产数据通常拥有长时间的历史记录, 这些数据可能使用不同的标准和格式, 遗留数据的存在使得数据治理过程中需要考虑如何与这些数据互操作, 以确保完整性和一致性^[4]。油田生产数据来自多个油井、设备和传感器, 这些数据源可能使用不同的硬件和软件系统, 以及不同的数据采集协议, 这多源性增加了数据标准化的挑战。油田生产行业可能存在多个标准和规范, 不同公司和地区可能采用不同的行业标准, 这种标准不一致性导致了数据共享问题。由于油田生产数据通常包含敏感信息, 企业对数据安全和隐私保护持谨慎态度, 这可能导致企业不愿意共享数据, 除非存在严格的数据安全措施, 不同数据源的数据质量可能不一致, 包括数据准确性、完整性和

一致性,在共享数据之前,需要解决这些质量问题。数据标准化和共享需要适当的技术支持,包括数据转换工具、集成平台和数据交换协议,企业文化和组织结构可能不支持数据共享,部分企业可能存在数据孤岛,不愿意共享数据,因为数据被视为竞争优势^[5]。

2.2 数据处理技术落后

部分油田生产企业可能在数据处理技术方面缺乏最新的知识和技能,他们可能仍在使用过时的数据处理工具和方法,而没有跟上技术的最新发展,一些小型或中小型油田生产企业可能拥有有限的信息技术资源和预算,难以进行数据处理技术的升级和更新。企业文化和员工的数据处理意识可能滞后,员工可能对新技术持怀疑态度,或者不愿意接受培训以学习新的数据处理方法,油田生产数据通常非常复杂,包括多源数据、多种数据类型和大规模数据。处理这些数据需要高级的技术和工具,而不是简单的解决方案,部分油田生产企业可能受到预算限制,难以购买和维护最新的数据处理软件和硬件。

2.3 专业人才匮乏

油田生产数据治理需要高度专业化的技能,包括地质学、工程学、计算机科学、数据科学等领域的知识,这些领域的专业人才通常需要多年的培训和经验积累,数据治理技术和工具不断发展和演进,需要专业人才不断学习和更新知识,缺乏持续学习和专业发展的机会可能导致专业人才匮乏。在油田生产数据治理领域,高水平的专业人才非常有竞争力,他们通常受到多家公司的争夺,这导致了供需不平衡,使得企业难以招聘到符合要求的专业人才,有些地区可能缺乏油田生产数据治理领域的专业人才,这可能受地理位置和行业基础设施的限制。高级的数据科学家、工程师和专业人才通常可以获得高薪,这导致企业难以支付他们的薪水,尤其是小型或中小型企业,油田生产数据治理领域的专业人才需要相对较长的培养周期,包括本科和研究生教育以及实际工作经验,这使得专业人才供应增长较慢。

3 油田生产数据治理技术及应用研究

3.1 数据治理流程

收集来自各种传感器、设备和系统的生产数据,包括地质勘探数据、钻井数据、生产数据、设备传感器数据等。确保数据的实时性和准确性,进行数据清洗,包括去除重复数据、处理缺失值、纠正错误数据等,对数据进行预处理,以确保数据的质量和可用性。

将处理后的数据存储存储在合适的数据仓库或数据库中,选择合适的存储技术和架构,以满足数据规模和性能要求,标准化数据,以确保数据格式和结构的一致性,这有助于不同数据源的数据互操作和集成。制定数据安全策略,以确保敏感数据的保护和访问控制,确保遵守法规和标准规范要求。运用数据分析和数据挖掘技术,揭示数据中的趋势、模式和关键信息,这有助于做出更好的决策和优化生产过程,使用可视化工具和仪表板,将数据呈现给决策者和非技术人员,以便他们更好地理解数据,生成定制化的报告和分析结果。制定数据备份和恢复计划,以确保数据可持续性和防止数据丢失,管理数据的生命周期,包括数据的存储、检索、维护,确保数据保持高质量和可用性。部署自动化和智能化技术,以监测设备状态、检测异常情况和自动化决策,这有助于提高生产效率和降低风险。

3.2 数据治理技术

在进行生产数据清洗的过程中,数据去重技术通过对数据进行比较和识别重复记录,将重复的数据删除,从而减少数据冗余,可以减少存储空间占用,简化数据分析和查询,降低数据处理成本,但是需要计算资源来查找和比较重复数据,可能导致一些误删。数据规范化技术通过将不同格式的数据统一到一个标准格式,以确保数据一致性,其可以提高数据一致性,简化数据集成和分析,但是可能需要复杂的转换规则,消耗计算资源。异常值检测技术用于识别数据中的异常或不寻常值,以确定是否需要对其进行清洗或进一步调查,其可以提高数据质量,减少错误数据的影响,增加数据可信度,但是需要选择合适的异常值检测算法,可能需要人工干预。数据填充技术用于填补数据中的缺失值,通常通过统计方法或插补方法完成,其可以维护数据的完整性,允许更多数据分析,但是填充可能引入伪造数据,影响数据准确性。

在进行数据挖掘的过程中,聚类分析通过将数据点分成不同的组,使每个组内的数据点相似,而不同组之间的数据点不相似,它依赖于数据点之间的距离度量,其可以发现数据中的隐藏模式,帮助识别相似性和差异性,可用于资源优化和设备故障检测。分类分析用于将数据点分为不同的类别,根据已知的类别标签,构建分类模型,以对新数据点进行分类,可以用于预测、识别和分类,支持决策制定和异常检测。关联规则挖掘用于发现数据中的项集之间的关联性,以确定项集之间的频繁出现关系,其可以用于发现数

据中的关联性。回归分析用于建立数据点之间的关系模型,以预测一个或多个目标变量的值,基于已知的自变量。

在对生产数据进行大数据分析的过程中,分布式计算技术是一个开源分布式存储和处理框架,它通过将大数据分割成小块,分布式存储和处理这些块,以实现高性能的数据处理,其具有高度可扩展性,容忍节点故障,适用于大规模数据分析。Spark 技术是一个快速的数据处理框架,通过内存计算和分布式计算模型,提供高性能的数据分析,其具有高性能、易于使用、支持多种编程语言的应用优势,适用于迭代式算法。机器学习和深度学习算法用于训练模型,从数据中识别模式和进行预测,以支持数据分析和决策制定。

在保障数据安全方面,加密技术使用密码学方法,将数据转换为密文,以保护数据的机密性,只有具有正确密钥的用户才能解密并访问数据,其提供强大的机密性,有效保护数据免受未经授权的访问。访问控制技术基于身份验证和授权原则,确保只有经过授权的用户能够访问特定数据或系统资源,其可以精确控制数据访问,降低数据泄漏风险。网络安全技术包括防火墙、入侵检测系统、虚拟专用网络等,用于保护数据在网络传输过程中的安全性。

3.3 数据治理要点

在开展油田生产数据治理的过程中,首先需要做好元数据管理工作,可以创建一个集中的元数据存储库,用于记录和管理与油田生产数据相关的元数据,这个存储库可以是数据库、元数据管理工具或文档库。确保一致性和标准化,制定元数据标准,包括数据字段命名规范、数据分类体系和元数据记录格式,为每个数据集或数据源创建元数据文档,其中包括数据的定义、来源、结构、格式、质量指标、更新频率和关联文档,建立数据字典,记录数据字段的定义、数据类型、单位、取值范围和数据元素之间的关系,数据字典应该是易于查阅和维护的。记录数据流程和数据映射,以识别数据的流动路径和变换过程,这有助于理解数据的整个生命周期,将数据分类,根据敏感度和重要性分为不同等级,并进行相应的安全标记。跟踪数据版本,确保能够回溯到以前的数据状态,这对于数据的可追溯性非常重要,建立元数据搜索和检索机制,使用户能够轻松查找和访问元数据,这可以通过元数据管理工具或搜索引擎实现。

同时,确保数据的质量是数据治理的核心要点,数据应准确、一致、完整、可信、可用和及时,数据质量评估、数据清洗和数据标准化是实现高质量数据的关键步骤,保护数据的安全性和隐私是至关重要的,采取适当的安全措施,包括加密、访问控制、网络安全和身份验证,以防止未经授权的访问和数据泄漏。建立明确的数据访问控制和权限管理策略,确保只有经过授权的用户能够访问敏感数据。审计数据访问,及时撤销不必要的权限,对数据进行分类,根据敏感程度和重要性将数据分为不同等级,不同等级的数据需要不同的安全措施和访问控制。建立清晰的数据治理策略和政策,包括数据所有权、数据流程、数据质量标准 and 数据治理团队的职责,组建跨职能的数据治理团队,包括数据管理员、数据质量负责人和数据用户,以协同管理数据,数据治理是一个持续改进的过程,定期审查数据治理策略、数据质量和安全性,识别和解决问题,以适应不断变化的需求。定期进行风险评估,识别数据管理中的潜在风险,并采取措施降低风险。

4 结论

综上所述,在油田开展生产作业的过程中,将会产生大量的生产数据,这些生产数据对于油田的生产安全 and 生产效率十分关键,由于数据量相对较大,因此,需要开展数据治理工作,目前在开展数据治理建设的过程中仍然存在多种类型的问题,需要采取合理的数据治理技术和数据治理流程,把控数据治理的要点,进而使得数据治理质量可以全面提升。

参考文献:

- [1] 甘振维,路智勇,李四海,等.江汉油田数字化转型探索与实践[J].石油科技论坛,2023,42(03):41-47.
- [2] 段鸿杰,马承杰,董琰,等.胜利油田数字化转型实践与成效[J].石油科技论坛,2023,42(03):48-55+64.
- [3] 张新.智能油田的数据治理工程及应用技术研究[J].信息系统工程,2023(05):52-54.
- [4] 邓红梅,姚卫华,焦扬,等.基于油气田企业的数据治理方法研究[J].价值工程,2023,42(05):11-13.
- [5] 马华锋,马荣,高鹏,等.数据治理在石油企业中的应用实践研究[J].设备管理与维修,2022(22):1-2.

作者简介:

王楠(1991-),女,辽宁鞍山人,研究生,工程师,主要从事油田数智化领域的技术研发工作。