

大数据技术在铁路货运平台中的运用探析

张志强

国能朔黄铁路肃宁分公司神港站

DOI: 10.12238/ems.v7i2.11676

[摘要] 随着中国国家铁路集团货运组织改革的持续深化,传统数据库技术在处理货运业务数据时捉襟见肘的困境,无法满足日益增长的数据分析与决策支持需求。本文聚焦于大数据技术在铁路货运平台中的运用,系统分析了前大数据技术在铁路货运领域的应用现状,提出了一套铁路货运大数据平台的构建方案,详细阐述了平台的实施步骤、架构设计以及关键技术应用,旨在通过大数据技术深入挖掘货运数据的隐藏价值,为铁路货运业务的智能化、高效化转型提供有力支撑。

[关键词] 大数据技术; 铁路货运平台; 数据挖掘

引言

随着信息技术的飞速发展,大数据技术已成为推动各行各业变革的重要力量。在铁路货运领域,传统的运行管理方式存在着系统独立和数据耦合程度不高、数据来源不准确等问题,迫切需要利用大数据进行创新。大数据是指通过采集、处理、分析大量数据,为铁路运输平台提供准确实时的信息支撑,进而达到智能、高效运行的目的。基于此,本文旨在对大数据技术在铁路货运平台上的应用进行深入研究,希望能对我国铁路货运运输产业的转型与升级起到一定的借鉴作用。

1 铁路货运大数据现状

1.1 系统独立性与数据耦合度问题

近几年来,许多业务系统已经在铁路货运业中得到推广和应用,例如货物电子商务系统、东北快运系统、货票系统和货运站系统等。但是,虽然这几个系统在各自的行业中都扮演着非常重要的角色,但是他们之间的耦合程度比较低,而且大部分都是封闭的,没有进行跨系统的横向比较和综合分析,不利于对数据的深度开发和利用,导致数据资源不能最大程度的发挥。

1.2 数据分析技术能力的局限

当前铁路货运数据大部分分析工作仍停留在传统的手工统计或数据库技术层面,这在处理小规模数据时或许尚能胜任,但在面对大数据的汹涌浪潮时却显得力不从心。然而,由于数据的可用性不足,即数据的质量、完整性和准确性都

很难得到保障,从而导致数据的可靠性和有效性,而数据的时效性又很差不能及时地反映市场的动态和变化的市场需求,这就限制了铁路货运大数据的价值得到很好的发挥,大大降低数据分析对商业决策的支撑作用^[1]。

2 大数据技术在铁路货运平台中的运用

2.1 平台数据采集层

铁路的内部数据主要来源于货运电商和快运等系统,而其业务的发展也需要对外界的信息如政府规划、客户需求、经济政策等进行大量的搜集,这都会对铁路货物运输产生很大的影响。同时,铁路运输企业也要密切关注公路、水路运输等行业竞争者的定价策略和相应的运量信息,才能在市场上保持竞争力。针对这一问题,货物运输大数据平台使用可以模仿人类的浏览方式,自动从万维网中获取所需要的信息的网络爬行技术。为了提高数据抓取效率和性能,平台进一步采用Java多线程技术,通过并行处理多个URL连接,Java多线程技术能够充分利用服务器的计算资源,实现快速且高效的信息采集。此外,为了降低服务器的内存消耗,平台还采用广度优先策略进行网页抓取,以分层的方式进行,首先抓取起始网页上的所有链接,然后再从这些链接指向的网页中抓取新的链接,以此类推。

2.2 平台数据传输层

在铁路货运大数据平台体系结构中,数据传输层覆盖了作为内部数据交换的主干网络,同时也包括与外部互联网的数据交互。为保证数据在内网和外网间的安全传递,必须构

建一道将内外部网络环境隔离开来的坚固壁垒, 保证互联网数据向内部网传递的安全。此外, 集团总部与各货运站段之间的数据交换需求, 则依托高效稳定的铁路办公网作为网络载体得以满足^[2]。

2.3 平台数据存储层

数据存储层则是铁路货物运输大数据平台的核心组成部分, 担负着维护和管理数据的重任。对于格式数据, 该平台使用行业中最先进的 Oracle 数据库来存储。Oracle 数据库具有数据处理能力强、可用性高、可扩充性好等特点能够满足铁路货物运输的需要。但是, 在铁路货物运输过程中, 不仅存在着海量的格式数据, 同时也存在着非结构化的信息, 如音频、视频等。这类数据由于其特殊性, 传统的关系型数据库难以满足其存储需求, 因此平台需要采用分布式文件系统(HDFS)作为非结构化数据的存储方案。HDFS 以其高容错性、高吞吐量的特点, 为海量非结构化数据的存储提供强有力的支持。对于从互联网获取的数据, 平台则通过专业的数据转换工具(如 Sqoop 等)将其转化为文件形式进行存储。

2.4 平台数据分析层

数据分析层是对各种信息源进行深度挖掘并对其进行分析, 针对结构化和非结构化数据的特点, 对其进行有针对性的处理。针对结构化数据(例如, 电子运单等), 其数据分析层主要依靠关系数据库来实现对其进行有效的存储和分析。其中, 数据仓库的联机分析处理(OLAP)是一种重要的研究手段, 不仅可以对数据进行快速的检索和多维度的分析, 而且可以挖掘出数据间的内在关联, 为企业的经营管理提供强有力的支撑。然而, 在铁路货运业务中, 同样存在着大量的非结构化数据, 如音频、视频等多媒体资料, 这类数据由于其非标准性和复杂性, 难以直接应用关系型数据库进行存储与分析, 因此数据分析层采用 NoSQL 数据库作为非结构化数据的存储容器, 并结合 Hadoop、Spark 等大数据处理框架进行深度挖掘。Hadoop 框架在处理低层次的分布式数据时利用 Hadoop 分布式文件系统(HDFS)实现数据的可靠存储与高效管理, 同时提供丰富的数据处理工具能够应对大规模数据集的复杂计算任务。通过 Hadoop, 数据分析层能够轻松处理那

些以音频、视频形式存在的非结构化数据, 为后续的深度分析奠定基础。与 Hadoop 相比, Spark 架构更注重机器学习、批处理等高层数据分析, 具有较 Hadoop 更为灵活、高效的数据处理机制, 可以快速应对复杂数据分析的需要。在此基础上, 采用回归分析、统计分析和神经网络等方法, 对数据进行深度挖掘, 为实现铁路货物运输管理的优化和创新提供技术支持^[3]。

2.5 平台数据展示层

在铁路货运大数据平台体系结构中, 数据表示层的关键问题是如何将数据快速准确地呈现给决策者, 使其能快速掌握商业动态并作出智能决策。为达到这个目的, 在平台上的数据呈现层通常会使用一种专门的工具——润乾报告, 它具有很强的定制功能和数据分析功能, 可以针对客户的具体需要, 来设计出既复杂又精确的报告。除了润乾报表外, 平台还引入 ECharts 这一基于 html5Canvas 的图表库, ECharts 具有良好的可扩展性、可视化的特性, 可为用户提供直观、形象、可互动的数据可视化图形, 帮助使用者更好地了解数据, 并从中挖掘出其中的规律与趋势。同时, ECharts 支持多种可视化类型, 包括散点图、折线图、柱状图、饼图、雷达图等, 能够满足不同场景下的数据展示需求。

结论:

因此, 将大数据技术应用于铁路货物运输平台是一项非常有意义的工作。通过运用大数据, 铁路货运平台在提高生产效率优化营销策略的同时, 也强化了企业的管理, 达到智能化、高效率发展的目的。在今后, 通过对大数据的深入研究可以更加深入地发掘大数据的价值, 促进产业创新与提升服务于经济与社会的发展。

[参考文献]

- [1]江月. 探讨大数据技术在电子商务数据分析中的应用[J]. 中国储运, 2024, (04): 109-110.
- [2]高文博. 大数据技术在计算机网络信息管理中的应用研究[J]. 无线互联科技, 2021, 18 (11): 85-86.
- [3]张明. 大数据技术在铁路运输设备技术状态管理中的应用[J]. 电子元件与信息技术, 2020, 4 (08): 72-74.