

基于工业互联网背景的大数据平台建设研究

许基伟, 马 欣

(南京林业大学土木工程学院, 江苏 南京 210000)

摘 要:随着信息技术和工业技术的迅猛发展,新型工业化道路应运而生,而工业大数据的平台建设对加快新型工业化道路具有十分重要的意义,本文通过对工业和互联网大数据进行对比分析,结合工业互联网与大数据的特点,阐述了二者之间的内在联系。基于此,提出了工业大数据平台架构以及相关的数据分级处理流程,最后指出了工业大数据平台的应用,包括资产管理、数据管理、数据分析和安全服务,为工业企业提供了平台支持和数据服务。

关键词: 工业互联网; 大数据; 平台建设

中图分类号: TP392 **文献标识码:** A

Research on the Construction of Big Data Platform Based on Industrial Internet

XU Jiwei, MA Xin

(College of Civil Engineering, Nanjing Forestry University, Nanjing 210000, China)

Abstract:With the rapid development of information technology and industrial technology, a new road for industrialization has come into being and industrial big data platform construction is very important for speeding up the development to industrialization. Through comparative analysis of industry and big data on the Internet, the paper combines the characteristics of industrial Internet and big data to elaborate on the inner link between them. Accordingly, it puts forward the major industrial big data platform architecture and related data classification processing flow, and finally points out the application of industrial big data platform, including asset management, data management, data analysis and security services, which provides platform support and data service for industrial enterprises.

Keywords: industrial Internet; big data; platform construction

1 引言(Introduction)

随着信息化和产业化进程的加快,新型工业化道路应运而生,以运用现代信息技术,用信息化带动工业化。工业4.0和中国制造2025都以创新为共同发展理念,强调以信息技术为载体,加强创新驱动与智能化服务水平。因此强调建设工业化的网络平台,对数据进行整理、分析、总结,工业企业应当发挥工业互联网的作用,形成终端的服务信息系统。通过建设大数据的平台,为传统工业化的改造指明了方向,对新型工业化建设具有重要意义^[1]。对工业互联网信息进行智能处理,实现数据同步、交换、集成、调用等功能,为工业企业提供数据平台服务和决策支持^[2]。

2 工业大数据分析与应用(Analysis and application of industrial big data)

2.1 工业大数据分析

大数据特征体现在量、速度、多样性、真实性四个方面,而工业大数据的特征在结合大数据的基础上,附加了可见性和价值两个特点。工业大数据历经了三个阶段,如表1所示。工业大数据与互联网大数据相比,最大的区别在于工业

大数据具有很强的目的性,而互联网大数据更多的是一种关联的挖掘,是更加发散的一种分析,对数据的预测和解读显得尤为重要^[3],如表2所示。

表1 工业大数据分析及应用三个阶段

Tab.1 Three stages in the analysis and application of industrial big data

时间	第一阶段	第二阶段	第三阶段
	1990—2000	2000—2010	2010—至今
核心技术	远程操作,实时更新管理	数据分析,中心平台数据支撑	数据分析平台与高级数据分析工具
问题对象/价值	对产品实行动态监控,实时进行故障检测和管理,一旦发现问题及时处理,以减少损失	以平台中心的大数据支撑,分析数据并及时提供线下服务措施,确保产品的顺利生产	用户为最高受益者,以高级数据分析软件为基础,实现以用户为核心的产品导向一站式服务
商业模式	机器式的操作服务	租赁体系和产品服务合同	按需的个性化自服务模式,分享经济
代表性企业和技术产品	GM OnStarTM	阿尔斯通Track TracerTM	GE Predix平台

表2 互联网大数据与工业大数据的对比分析

Tab.2 Comparative analysis of Internet big data and industrial big data

属性	互联网大数据	工业大数据
数据量需求	较多但不详细	较多且详细
数据质量要求	较低	较高
数据的属性意义	只针对数据本身，不考虑其关联性	重视数据本身的物理关联和相关性
准确性要求	较低，仅针对事物本身分析	较高，注重跨学科的技术分析和数据管理

2.2 工业大数据应用

随着信息技术的迅猛发展，工业企业也相继进入了互联网工业的新的发展阶段，工业大数据在此背景下创新和变革，其应用范围很宽泛。工业大数据通常应用在制造、航空、轨道交通、船舶、石油、建筑等方面^[4,5]，如表3所示。

表3 工业大数据的应用

Tab.3 Application of industrial big data

领域	应用
制造	绿色工厂、智慧物流产品服务、智能产品
航空	智能航行线路、安全管理、机器故障排查
轨道交通	列车运维、行车安全、环境安全
船舶	设备综合保障、降低船舶能耗、航海安全数据服务
石油	勘测数据管理和共享、油气生产物联网、管道完整性管理
建筑	绿色建筑、新型材料、规划设计、旧城改造

3 工业互联网与大数据(Industrial Internet and big data)

3.1 工业互联网与大数据的联系

工业企业发展的动力来自工业互联网与大数据的融合，通过工业互联网，将来自数据操控平台中的信息汇总，依据相应的产品技术要求，实现数据的解读与分析，从而提炼出对企业有价值的信息。而大数据可以在跨学科技术融合的基础上，进行信息汇总与管理，提炼产品最具有价值的信息。通过互联网与线下的其他传播渠道，从而建立新投资目标、发现新趋势、提供解决复杂问题的新路径。

3.2 工业互联网与大数据的作用

工业互联网是互联网与智能制造的交叉点，在融合大数据的前提下，提升了产品的智能性，并充分拓展了行业的相关应用。产品的智能化是把机器处理和导出系统渗透到产品的生产过程中，保证产品的感知、存储等能力，实现产品的信息化定位、识别、复原。目前互联网汽车、工程机械、智能家电等是产品智能化的热点领域。同时，工业互联网和大数据通过网络连接到企业管理平台。企业管理平台可以利用无线网络、视频远程故障诊断等信息服务系统对设备运行实现远程操控，及时播报预警^[6]。

3.3 工业互联网与大数据的特点

在工业互联网时代，通过纳入来自产业链上下游以及跨

界的数据，实现工业从数据向大数据的转变。工业互联网与大数据主要呈现四个方面的特点，即：全要素、全过程、全方位、全融合。全要素是指为保证产品数据的完整性，其携带了产品全部的尺寸、工艺、制造、售后使用信息等；全过程是指为确保产品的最终质量，在数据设计和使用，时必须考虑跨越不同的设计、制造阶段；全方位指具体到产品的生产线，从设计、制造、采购等，全方位地关注相关信息，确保产品的实用性。全融合指在信息技术支持的背景下，关注企业各业务的全面关联及融合。

4 工业大数据平台建设(Construction of industrial big data platform)

4.1 云平台总体架构

工业大数据云平台通过寻找统一数据源、数据口径和数据出入口，来实现对各种专业的合理分析，并最终统一和完善企业信息模型。云平台的总体架构包括PaaS环境层、PaaS业务层、PaaS服务层。PaaS环境层为业务应用提供支撑的软件组件、各种中间件和数据库等，以Hadoop为代表的大数据处理；PaaS业务层包含了应用的后台程序、数据处理算法以及业务数据等实现业务能力的元素；PaaS服务层是指将业务层的业务、算法和数据以接口的形式提供给上层的前端应用直接访问。总体来说，云平台总体架构面向一般数据中心典型的应用场景，提供对混合IT资源的统一接入，以构筑云模式下基础资源调度的最佳实践，同时以PaaS能力为核心，将应用系统的典型软件组件以服务形态提供，为业务系统提供统一环境支持，并进行统一管理和监控，将大数据平台作为典型服务组件整合到云平台中进行统一管理，以适应未来应用对大数据能力的普遍使用。为用户更好地提供面向DevOps的统一云服务业务流程，以统一平台提供传统的IaaS和PaaS能力，并贯穿开发、测试和生产的全过程^[7]。

4.2 大数据平台目标架构

大数据平台目标架构包括五个层面：应用层、能力层、数据层、获取层、数据源。如图1所示。其中Hadoop平台是大数据平台中的核心。通过建立分布式的文件管理系统，保证了大量数据安全存储，建立了数据分析处理框架。Hadoop集群能迅速扩展到顶级服务器，实现对大数据的批量处理。同时，Hadoop通过MapReduce可将任务分布并行运行在一个集群服务器中，实现大规模并行计算，同时考虑到设备的不稳定性，保证了计算的准确性和高效性。而Hadoop云平台以半结构化数据和非结构化数据为主，通过存储海量数据，确保使用时的时效性与高效性。

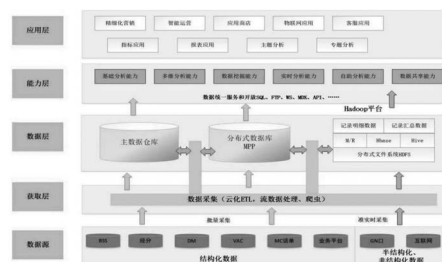


图1 大数据平台目标架构

Fig.1 Target architecture of big data platform

4.3 大数据平台中的数据分级存储

数据分级存储应当满足一定的原则,即大数据平台强调数据的生命周期性,随着数据生命周期的变化逐步向通用性能存储迁移,是分级存储管理的主线。同时数据分级存储在保证主线畅通的情况下,考虑到其他分级原则,确保数据迁移能够覆盖到各个层面。在满足相关原则的基础上,将核心模型改造,转化为现有主数据仓库的核心模型,减少了数据冗余。集成改造后,将主数据仓库中的关键性数据转移到一个低成本的分分布式数据库中,降低了主数据库的存储压力。同时此存储系统也支持数据的深度分析。

4.4 大数据平台中的数据处理流程

在满足相关技术的基础上,数据的采集工作通常包括结构化和非结构化的两种数据采集,数据的具体处理过程分为七个环节:①源数据导入ETL,进行数据的清洗、转换和入库;②基础数据加载到主数据仓库,规划保存三年;③清洗、转换后的ODS加载到分布式数据库规划保存两个月,具体数据分析和汇总在分布数据库中进行,规划保存两年;④ODS数据和非结构化数据,如爬到的网页数据ftp到Hadoop平台做长久保存;⑤非结构化数据分析处理在Hadoop平台完成,产生的结果加载到分布式数据库;⑥生成KPI和高度汇总数据加载到主数据仓库;⑦业务应用通过数据访问接口获取所需数据。

5 工业大数据平台应用(Application of industrial big data platform)

5.1 资产管理服务

资产管理服务的重点目标是资产建模、连接资产和数据源,其服务范围分为三个板块:接口层支持资产分层、属性分类和自定义建模对象,终端访问对象建模层,通过转换数据格式,以便于数据库存储和查询;相关查询引擎利用图形表达式语言检索数据;图形数据库中包含资产服务数据存储,并将原始数据描述成数据模型。资产管理服务中的资产模型是工业大数据平台的关键,是连接平台所有服务的中枢,能帮助应用开发者更好的理解、分析和处理数据。

5.2 数据管理与服务

大数据平台常用的数据管理与服务包括对各种类型数据进行接入、处理和储存。数据接入指支持多渠道、多类型的数据接入,包括实时数据接入和批量数据接入;数据处理包含对传感器数据进行标签处理、资产数据与ERP数据相结合等;数据储存是根据不同的需求选择相应的存储方式,类似时间序列存储传感器实时数据、BLOB数据库存储图片式数据、关系型数据库存储其他数据等。

5.3 数据分析服务

数据分析服务通常指通过分析实时数据检测设备状态、预防设备故障、优化生产过程等,并进行迭代提升,在历史数据进行整合和分析,包括异常检测、事件流处理、分析运行环境、分析界面、分析日志等服务项目,建立工业级的预测模型,以进行更有效地生产和运营。相较于数据自身的

特点,数据分析服务在涉及技术层面的基础上,更应当满足企业与客户的应用需求,将极度专业化的技术描述,转化为更容易的商业需求。

5.4 数据安全服务

数据安全服务的主要目标为建立企业互联网应用,解决安全性问题,从而满足企业互联网端对端的安全需求。其服务范围包括租户管理、用户账户和身份验证、访问控制三个方面。租户管理指为租户提供特定服务实体及注册表,完成租户和服务实体间的完美映射,在确保客户数据信息安全的基础上,实现服务生命周期管理。用户账户和身份验证是在传统用户信息验证方法的基础上,增加多种验证方法完善身份验证系统。访问控制是指建立资源访问限制,增强用户账户信息验证的权威性,同时不断提升和优化网络访问信息安全,实现复杂条件下的访问控制。

6 结论(Conclusion)

随着新型工业化道路的推进,传统的工业化道路已不能适应当前社会的发展需求,本文在结合信息化和工业化的手段,搭建了工业大数据平台,并应用于资产管理、数据管理、数据分析、数据安全服务等方面,未来工业发展的趋势应当在构建平台的基础上,实现产品智能化,确保产品生产的高效性。同时工业大数据平台也为工业企业提供了商业支持和服务。

参考文献(References)

- [1] Du B,Huang R,Chen X,et al.Active CTDaaS:A Data Service Framework Based on Transparent IoD in City Traffic[J].IEEE Transactions on Computers,2016,65:1-10.
- [2] Valerio Persico,Antonio Pescapé,Antonio Picariello,et al. Benchmarking big data architectures for social networks data processing using public cloud platforms[J].Future Generation Computer Systems,2018,89:98-109.
- [3] Jeon S,Hong B.Monte Carlo simulation-based traffic speed forecasting using historical big data[J].Future Generation Computer Systems,2016,65:182-195.
- [4] 刘德贺,史玉峰.基于TLS技术的道路边坡地表位移监测研究[J].森林工程,2018,34(02):95-99.
- [5] 吴佳蔚,李希胜,曹璐琳,等.基于BIM技术的绿色建筑节地设计自动评价[J].建筑科学,2017,33(04):123-128.
- [6] 郑加柱,高业何敏.自动推算室内接入点坐标算法[J].测绘通报,2018(02):21-24;29.
- [7] 魏洋,卢海鹏,刘小明,等.基于长标距光纤光栅传感器的连续梁桥监测技术[J].铁道科学与工程学报,2017,14(10):2231-2238.

作者简介:

许基伟(1996-),男,硕士生.研究领域:数据平台.

马欣(1977-),男,博士,副教授.研究领域:房地产经济与数据平台搭建.