

基于物联网的能源企业大数据存储技术研究

孔超民¹ 孔维洋²

(1. 山东深数信息科技有限公司 济南 250014;

2. 济南柏腾电子科技有限公司 济南 250014)

摘要 针对能源企业在大数据存储技术方面存在的高延迟、数据处理效率低等问题,文中提出了一种基于物联网边缘计算的解决方案。该方案旨在通过在数据源附近实施边缘计算,优化能源企业的大数据存储和处理流程。该方法将数据处理任务从远程云中心转移到网络边缘,不仅降低了数据传输延迟,还提高了数据处理的效率。实验结果显示,相较于传统的中心化数据存储方法,该方案在处理能源数据时表现出更高的效率和更快的响应速度。该结果为能源企业提供了一种高效的大数据存储技术方案,也有助于优化能源数据的管理和分析流程,同时提升整体数据处理性能。

关键词: 能源企业大数据存储;物联网技术;边缘计算技术

中图分类号 TN929.5

Research on Big Data Storage Technology of Energy Enterprise Based on Internet of Things

KONG Chaomin¹ and KONG Weiyang²

(1. Shandong Shenshu Information Technology Co., Ltd., Jinan 250014, China;

2. Jinan Boteng Electronic Technology Co., Ltd., Jinan 250014, China)

Abstract Aiming at the problems of high latency and low data processing efficiency in big data storage technology in energy enterprises, this paper proposes a solution based on Internet of Things edge computing. The scheme aims to optimize the big data storage and processing flow of energy enterprises by implementing edge computing near the data source. The method shifts the data processing task from the remote cloud center to the network edge, which not only reduces the data transmission delay, but also improves the efficiency of data processing. The experimental results show that compared with the traditional centralized data storage method, the scheme shows higher efficiency and faster response speed when processing energy data. The results provide an efficient big data storage technology solution for energy enterprises, which also helps to optimize the management and analysis process of energy data, and improve the overall data processing performance.

Key words Energy enterprise big data storage, Internet of Things technology, Edge computing technology

0 引言

随着能源行业的快速发展与数字化转型,能源企业正面临着越来越多的数据处理挑战^[1]。在该背景下,大数据技术成为支持能源企业高效运营的关键工具^[2]。随着物联网技术的广泛应用,能源企业能通过各种传感器和智能设备实时收集关于能源生产、分配和消费的海量数据^[3]。然而,这些大规模的数据集带来了存储和处理的难题,尤其是在数据传输延迟、中心化处理效率等方面^[4]。传统的中心化数据存储方法往往难以满足实时数据分析和快速响应的需求,这在管理和决策过程中可能导致效率损失^[5]。

边缘计算作为一种新兴的技术,提供了一种可行的解决方案。边缘计算通过在数据产生的地点附近进行数据处理,减少了对远程数据中心的依赖,显著降低了数据传输延

迟。此外,边缘计算还可以提高数据处理的效率和安全性,因为它允许在本地快速分析和响应数据,减少了数据在传输过程中的潜在风险。

本文提出了一种基于物联网边缘计算的能源企业大数据存储和处理方法。通过将边缘计算技术应用于能源企业的大数据管理,探索提高数据处理效率、降低延迟、增强数据安全性的新途径,以期为能源行业的数字化转型提供支持。

1 相关技术基础

1.1 物联网技术

物联网技术是一种涵盖了传感器、嵌入式系统、通信协议、云计算等领域的综合性技术体系。其核心思想是通过

作者简介:孔超民(1988—),本科,研究方向为物联网与大数据。

连接和互联各种物理设备和对象,实现数据的采集、传输、处理和应用。在能源领域,物联网技术可用于实时监测能源设备的运行状态,收集能源消耗数据,并与能源系统进行远程通信。传感器网络和物联网协议的应用使能源企业能实时获取有关能源生产和分配的关键信息,从而优化运营和资源管理。

1.2 边缘计算技术

边缘计算技术是一种分布式计算模型,其旨在将计算和数据处理任务从传统的云数据中心转移到数据源附近的边缘设备或边缘节点。边缘计算强调在数据产生的地方进行数据处理,以降低数据传输延迟,提高数据处理效率,减轻中心数据中心的负担。在能源企业中,边缘计算技术可以应用于实时监测、数据预处理、异常检测、实时决策支持等方面。它有助于加速数据分析过程,减少对网络带宽的需求,并提高能源系统的响应速度。

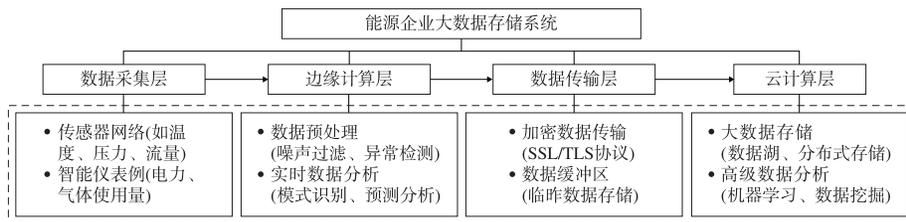


图1 基于边缘计算的能源企业大数据存储系统

整体而言,该系统架构充分利用了边缘计算的优势,即在数据源附近处理数据,以提高效率,降低延迟。同时,通过与云计算的结合,保证了数据存储的可靠性和高级数据处理的能力,为能源企业提供了一个全面、高效、安全的大数据存储和处理方案。

2.2 数据采集层

数据采集层的主要任务是通过传感器网络和智能仪表收集各类能源使用数据。这些数据包括温度、压力、流量、电力、气体使用量等。数据采集层的设计必须确保数据的精度和实时性,为后续的处理和分析提供基础。时间序列数据采集式如式(1)所示:

$$Y(t) = \alpha X(t) + \beta X(t - \Delta t) + \epsilon(t) \quad (1)$$

其中, $Y(t)$ 代表在时间 t 收集到的数据, $X(t)$ 是原始传感器读数, Δt 是时间间隔, α 和 β 是加权系数,用于调整当前读数与历史读数的相关性, $\epsilon(t)$ 是误差项,表示随机干扰。

接着,通过数据融合的方法整合来自多个传感器的数据,如式(2)所示:

$$Z = \sum_{i=1}^n w_i Y_i(t) \quad (2)$$

其中, Z 代表经过融合的综合数据, Y_i 是第 i 个传感器在时间 t 上的数据, w_i 是相应的权重系数,用于反映不同传感器数据的重要性和可靠性。

这样的处理过程不仅增强了数据的准确性和可用性,还为后续的边缘计算层提供了高质量的输入数据。

2 基于边缘计算的能源企业大数据存储技术

2.1 系统架构

本系统的设计理念是在数据的生产地附近对其进行初步处理和分析,以减少数据传输所需的带宽,提高数据处理的实时性,有效降低整体系统的能耗。该架构主要分为4个层次,即数据采集层、边缘计算层、数据传输层和云计算层。在数据采集层,系统利用传感器网络和智能仪表来收集原始数据。随后,在边缘计算层,这些原始数据会经过初步处理及实时数据分析,如模式识别和预测分析。数据传输层会使用加密数据传输技术确保数据在传输过程中的安全性。此外,数据缓冲区可以用于临时存储数据,以减轻网络高峰时段的压力。最终,在云计算层,大数据存储和高级数据分析的功能得以实现。系统架构如图1所示。

2.3 边缘计算层

边缘计算层的主要职责是对来自数据采集层的数据进行实时处理和初步分析。该层紧邻数据产生点,可利用边缘节点的计算能力快速进行数据预处理、异常检测、局部分析,从而降低数据对云端的需求和网络延迟。

该层首先会使用正态分布进行数据的异常检测和模式识别,如式(3)所示:

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

其中, $F(x)$ 表示正态分布函数, μ 和 σ 分别代表数据的均值和标准差。

接下来,需要评估数据的时间相关性,如预测维护和负荷预测。利用截断指数分布,边缘计算层能有效地处理对时间较为敏感的数据,优化能源分配和使用效率,如式(4)所示:

$$G(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda T}} \quad (4)$$

其中, $G(x)$ 是一个截断指数分布函数, λ 是率参数, T 是观测窗口。

通过边缘计算层,系统能显著提升整体的数据处理速度和效率,同时降低对中心化云资源的依赖。

2.4 数据传输层

数据传输层的核心任务是传输加密数据,同时保证数据在传输过程中不受未经授权的访问或篡改。为此,数据传输层通常采用高级加密标准和安全传输协议,以确保数

数据传输的安全性。同时,为提高传输效率,数据传输层还会采用数据压缩技术来控制数据的大小。

在本系统中,该层使用对称密钥加密,其数学表达式如式(5)所示:

$$C = E_k(D) \quad (5)$$

其中, C 表示加密后的数据, E_k 是以 k 为密钥的加密函数, D 是原始数据。

数据压缩的核心目标是减少数据的体积,以提高传输效率。压缩的基本概念如式(6)所示:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (6)$$

其中, $H(X)$ 代表数据集 X 的熵, n 是数据集 X 中不同消息的总数, x_i 是数据集中的第 i 个消息或符号, $p(x_i)$ 是消息 x_i 出现的概率。

数据传输层在保证数据传输的安全性和高效性的同时,也保证了数据传输过程的稳定性和可靠性。

2.5 云计算层

云计算主要有两个关键组成部分,即大数据存储和数据分析。大数据存储主要依赖于数据湖和分布式存储技术。数据湖允许存储大量的结构化和非结构化数据,而分布式存储系统则提供了数据的高可用性和弹性扩展能力。这些技术确保了数据的安全存储,同时提高了数据访问的效率。数据存储优化公式如式(7)所示:

$$S = \frac{D}{N} + \sum_{i=1}^N \text{Overhead}_i \quad (7)$$

其中, S 表示总存储需求, D 是原始数据大小, N 是数据副本数量, Overhead_i 是第 i 个副本的存储开销。

进行数据分析时,机器学习算法不仅能处理结构化数据,还能有效处理非结构化数据,如文本和图像。该层使用的机器学习算法是支持向量机,它被广泛应用于分类和回归任务,如式(8)所示:

$$\min_{w,b,\xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i \quad (8)$$

$$\text{subject to } y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$$

其中, \mathbf{w} 是超平面的法向量, b 是偏置项, ξ_i 是将输入数据 \mathbf{x}_i 映射到更高维空间的函数, y_i 是数据点 \mathbf{x}_i 的类别标签。

云计算层不仅提高了数据处理的效率和准确性,还为能源企业提供了基于数据的深入洞察,从而帮助他们做出更明智的决策。

3 实验设计与结果分析

3.1 实验环境设置

本文实验的硬件配置如下。服务器型号为Dell PowerEdge R740,配备Intel Xeon Gold 6230处理器和128GB RAM。在软件配置方面,使用了最新版本的VMware ESXi 6.7作为虚拟化平台,以便在物理服务器上创建和管理多个虚拟机。对于大

数据存储和处理,选择将Apache Hadoop 3.2和Apache Spark 2.4作为主要的数据处理框架,并安装Python 3.7,NumPy,Pandas和Scikit-learn。实验仿真操作在MATLAB R2020a和Cisco Packet Tracer 7.2上进行。MATLAB R2020a用于模拟和分析数据处理算法,Cisco Packet Tracer 7.2则用于模拟网络环境和数据传输过程。

3.2 实验结果及分析

为对比本文提出的基于边缘计算的能源企业大数据存储方法和传统的中心化系统之间的性能差异,在相同的环境配置下对两个系统进行了测试,结果如表1所列。

表1 实验结果

测试指标	边缘计算系统	传统中心化系统
数据处理延迟/ms	15	50
数据吞吐量/(条/秒)	1 000	500
数据存储效率/%	95	75
响应速度/ms	100	300

实验结果表明,相较于传统的中心化数据存储方法,基于边缘计算的能源企业大数据存储系统在处理能源数据时,不仅提高了效率,还大幅提升了响应速度。这种优化对于提升能源企业的运营效率和决策质量具有重要意义。

4 结语

本文深入分析了基于边缘计算的能源企业大数据存储系统,该系统通过创新的多层架构有效整合了数据采集、处理和存储的过程。需要注意的是,边缘计算层在降低数据处理延迟和提高实时性方面发挥了关键作用,而云计算层则通过高级数据分析技术(如机器学习),实现了深入的数据挖掘和模式识别。系统的综合设计不仅提升了数据处理的效率和准确性,还保证了数据传输的安全性。实验结果表明,与传统的中心化数据存储方法相比,本系统在能源数据处理方面具有显著优势,特别是在响应速度和数据吞吐量方面。因此,该系统为能源企业提供了一种高效、可靠的大数据存储和处理方案,这对优化能源行业的数据管理和决策制定具有重要价值。

参考文献

- [1] 张岩,胡林生.大数据分布式存储技术在中小型金融科技企业的应用与推广[J].中国管理信息化,2021,24(11):108-110.
- [2] 曹杰.浅谈企业数据云数据存储技术[J].中国新通信,2020,22(3):85.
- [3] 岳茹.云计算环境下的数据存储与管理技术研究[J].科技资讯,2023,21(21):29-32.
- [4] 杨晓娟.基于云计算技术的网络安全数据存储系统设计[J].无线互联科技,2023,20(13):159-161.
- [5] 衣娜.基于云计算技术的网络安全数据存储系统设计[J].数字技术与应用,2023,41(6):237-239.