



(12) 发明专利申请

(10) 申请公布号 CN 120218855 A

(43) 申请公布日 2025.06.27

(21) 申请号 202510323405.7

G16Y 10/20 (2020.01)

(22) 申请日 2025.03.18

(71) 申请人 中煤科工开采研究院有限公司

地址 101399 北京市顺义区中关村科技园
区顺义园临空二路1号

(72) 发明人 吕依濛 牟振栋

(74) 专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

专利代理师 白雪静

(51) Int. Cl.

G06Q 10/10 (2023.01)

G06Q 50/02 (2024.01)

G06N 3/08 (2023.01)

G06N 3/0464 (2023.01)

G06F 16/28 (2019.01)

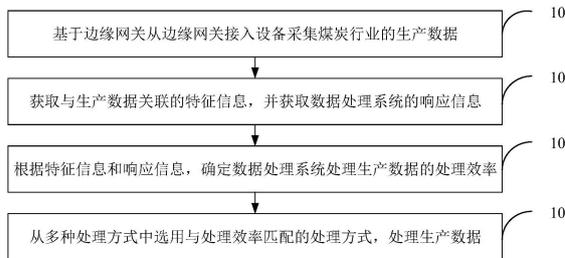
权利要求书2页 说明书10页 附图1页

(54) 发明名称

基于物联网技术的煤炭生产数据处理方法及系统

(57) 摘要

本申请提出一种基于物联网技术的煤炭生产数据处理方法及系统。其中,方法应用于数据处理系统,该方法包括:基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据;获取与生产数据关联的特征信息,并获取数据处理系统的响应信息;根据特征信息和响应信息,确定数据处理系统处理生产数据的处理效率;从多种处理方式中选用与处理效率匹配的处理方式,处理生产数据。本申请可以使得煤炭生产数据能够高效且精准地进行采集和计算,为煤炭行业的智能化生产、质量监控、环境监测等提供坚实的基础。



1. 一种基于物联网技术的煤炭生产数据处理方法,其特征在于,应用于数据处理系统,所述方法包括:

基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据;

获取与所述生产数据关联的特征信息,并获取所述数据处理系统的响应信息;

根据所述特征信息和所述响应信息,确定所述数据处理系统处理所述生产数据的处理效率;

从多种处理方式中选用与所述处理效率匹配的处理方式,处理所述生产数据;其中,所述多种处理方式包括流式数据处理方式、批量数据处理方式和混合处理方式,所述混合处理方式是指混合流式数据处理和批量数据处理的方式。

2. 如权利要求1所述的方法,其特征在于,所述根据所述特征信息和所述响应信息,确定所述数据处理系统处理所述生产数据的处理效率,包括:

根据所述特征信息和所述响应信息,确定多个实时处理效能信息;

根据所述多个实时处理效能信息,确定所述数据处理系统处理所述生产数据的处理效率。

3. 如权利要求2所述的方法,其特征在于,

所述特征信息包括以下至少一项:单位生产数据的处理时长、处理的单位生产数据数量、处理的生产数据总量、所述生产数据系统处理的复杂度、处理的生产数据最大延迟时间;

所述响应信息包括以下至少一项:系统处理数据的基准时长、系统处理单位生产数据的基准时长、系统实际延迟;

所述多个实时处理效能信息包括以下至少一项:实时处理率、实时接收率、实时处理数据量、实时处理复杂度、实时延迟率。

4. 如权利要求3所述的方法,其特征在于,所述根据所述特征信息和所述响应信息,确定多个实时处理效能信息,包括:

根据所述单位生产数据的处理时长和所述系统处理数据的基准时长,确定所述实时处理率;

根据所述处理的单位生产数据数量和所述系统处理单位生产数据的基准时长,确定所述实时接收率;

根据所述处理的生产数据总量和所述系统处理数据的基准时长,确定所述实时处理数据量;

根据所述生产数据系统处理的复杂度和所述系统处理单位生产数据的基准时长,确定所述实时处理复杂度;

根据所述处理的生产数据最大延迟时间和所述系统实际延迟,确定所述实时延迟率。

5. 如权利要求2-4中任一项所述的方法,其特征在于,所述根据所述多个实时处理效能信息,确定所述数据处理系统处理所述生产数据的处理效率,包括:

根据所述多个实时处理效能信息,采用第一公式计算所述数据处理系统处理所述生产数据的处理效率;其中,所述第一公式表示如下:

$$E = \frac{p1 * r1 + p2 * \arcsin(r2) + p3 * v1 * c1 + p4 * r3}{p1 + p2 + p3 + p4}$$

其中,E为所述数据处理系统处理所述生产数据的处理效率;r1为实时处理率,r2为实时接收率,v1为实时处理数据量,c1为实时处理复杂度,r3为实时延迟率,均为所述实时处理效能信息;p1、p2、p3、p4均为处理效能影响因子。

6.如权利要求1所述的方法,其特征在于,所述从多种处理方式中选用与所述处理效率匹配的处理方式,处理所述生产数据,包括:

所述处理效率小于第一阈值,采用与所述处理效率匹配的所述批量数据处理方式处理所述生产数据;或者,

所述处理效率大于第二阈值,采用与所述处理效率匹配的所述流式数据处理方式处理所述生产数据;或者,

所述处理效率大于或等于所述第一阈值且小于或等于所述第二阈值,采用与所述处理效率匹配的所述混合处理方式处理所述生产数据。

7.如权利要求6所述的方法,其特征在于,所述采用与所述处理效率匹配的所述混合处理方式处理所述生产数据,包括:

确定所述混合处理方式的数据分配率;

基于所述数据分配率,将所述生产数据拆分成第一部分生产数据和第二部分生产数据,其中,所述第一部分生产数据待用所述流式数据处理方式处理,所述第二部分生产数据待用所述批量数据处理方式处理。

8.如权利要求6所述的方法,其特征在于,

所述采用与所述处理效率匹配的所述批量数据处理方式处理所述生产数据,包括:

利用批量计算工具对所述生产数据进行数据加工与处理,并将处理完成的数据存储至数据仓库中;

所述采用与所述处理效率匹配的所述流式数据处理方式处理所述生产数据,包括:

将所述生产数据存储于Kafka集群中,并基于实时计算工具对所述Kafka集群中的生产数据进行数据运算与处理,将处理完成的数据存储至时序数据库;其中,所述时序数据库中的数据定期汇总至所述数据仓库中进行保存。

9.如权利要求1所述的方法,其特征在于,所述方法还包括:

在采用与所述处理效率匹配的处理方式处理所述生产数据的过程中,基于预设的预警规则判断是否触发报警事件,并在触发所述报警事件时,基于所述报警事件关联的报警方式进行报警。

10.一种数据处理系统,其特征在于,包括:

采集模块,用于基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据,并获取与所述生产数据关联的特征信息和所述数据处理系统的响应信息,根据所述特征信息和所述响应信息,确定所述数据处理系统处理所述生产数据的处理效率;

数据处理平台,用于从多种处理方式中选用与所述处理效率匹配的处理方式,处理所述生产数据;其中,所述多种处理方式包括流式数据处理方式、批量数据处理方式和混合处理方式,所述混合处理方式是指混合流式数据处理和批量数据处理的方式。

基于物联网技术的煤炭生产数据处理方法及系统

技术领域

[0001] 本申请涉及工业互联网技术、物联网与传感器技术、大数据技术和云计算等技术领域,尤其涉及一种基于物联网技术的煤炭生产数据处理方法及系统。

背景技术

[0002] 工业互联网是新一代信息通信技术与工业经济深度融合的新型基础设施、应用模式和工业生态。它通过对人、机、物、系统等的全面连接,构建起覆盖全产业链、全价值链的全新制造和服务体系,为工业乃至产业数字化、网络化、智能化发展提供了实现途径。在煤炭行业生产数据采集方面,工业互联网技术主要发挥着重要作用。

[0003] 开采生产数据的未来应用场景十分广阔,涵盖了智能工厂与自动化生产、供应链优化、能源管理、安全保障、产品质量监控与追溯、产品创新与优化以及销售预测与需求管等多个方面。但是,目前尚缺乏如何采集与处理煤炭生产数据的有效手段。

发明内容

[0004] 本申请实施例提供一种物联网技术的煤炭生产数据处理方法及系统。

[0005] 根据本申请实施例的第一方面,提供一种物联网技术的煤炭生产数据处理方法,应用于数据处理系统,上述方法包括:

[0006] 基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据;

[0007] 获取与生产数据关联的特征信息,并获取数据处理系统的响应信息;

[0008] 根据特征信息和响应信息,确定数据处理系统处理生产数据的处理效率;

[0009] 从多种处理方式中选用与处理效率匹配的处理方式,处理生产数据;其中,多种处理方式包括流式数据处理方式、批量数据处理方式和混合处理方式,混合处理方式是指混合流式数据处理和批量数据处理的方式。

[0010] 根据本申请实施例的第二方面,提供一种数据处理系统,包括:

[0011] 采集模块,用于基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据,并获取与生产数据关联的特征信息和数据处理系统的响应信息,根据特征信息和响应信息,确定数据处理系统处理生产数据的处理效率;

[0012] 数据处理平台,用于从多种处理方式中选用与处理效率匹配的处理方式,处理生产数据;其中,多种处理方式包括流式数据处理方式、批量数据处理方式和混合处理方式,混合处理方式是指混合流式数据处理和批量数据处理的方式。

[0013] 根据本申请实施例的第三方面,提供一种电子设备,包括:

[0014] 至少一个处理器;以及

[0015] 与所述至少一个处理器通信连接的存储器;其中,

[0016] 所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行前述第一方面所述的方法。

[0017] 根据本申请实施例的第四方面,提供一种存储介质,存储介质存储有指令,当指令

在电子设备上运行时,使得电子设备执行前述第一方面所述的方法。

[0018] 根据本申请实施例的第五方面,提供一种计算机程序产品,当计算机程序产品中的指令被处理器执行时,实现前述第一方面所述方法的步骤。

[0019] 根据本申请技术方案,可以根据采集的煤炭行业生产数据关联的特征信息和数据处理系统的响应数据,确定数据处理系统处理该生产数据的处理效率,便于选用与该处理效率匹配的处理方式来处理该生产数据,使得煤炭生产数据能够高效且精准地进行采集和计算,为煤炭行业的智能化生产、质量监控、环境监测等提供坚实的基础。

[0020] 本申请附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本申请的实践了解到。

附图说明

[0021] 本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0022] 图1为本申请实施例所提供的基于物联网技术的煤炭生产数据处理方法的流程图;

[0023] 图2为本申请实施例提供的数据处理系统的框图;

[0024] 图3为本申请实施例提供的数据处理系统的架构示意图。

具体实施方式

[0025] 下面详细描述本申请的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本申请,而不能理解为对本申请的限制。

[0026] 以下结合附图对本申请的示范性实施例做出说明,其中包括本申请实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本申请的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0027] 在本申请一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本申请一个或多个实施例。在本申请一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。还应当理解,本申请一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0028] 应当理解,尽管在本申请一个或多个实施例中可能采用术语第一、第二等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本申请一个或多个实施例范围的情况下,第一也可以被称为第二,类似地,第二也可以被称为第一。取决于语境,如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0029] 需要说明的是,在本申请的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。本申请所涉及的信息(包括但不限于用户设备信息、用户个人信息等)、数据(包括但不限于用于

分析的数据、存储的数据、展示的数据等)以及信号,均为经用户授权或者经过各方充分授权的,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

[0030] 需要说明的是,在本申请实施例中,可能提及某些软件、组件、模型等业界已有方案,应当将它们认为是示范性的,其目的仅仅是为了说明本申请技术方案实施中的可行性,但并不意味着申请人已经或者必然用到了该方案。

[0031] 下面参考附图描述本申请实施例的基于物联网技术的煤炭生产数据处理方法及系统。

[0032] 其中,需要说明的是,本申请实施例的基于物联网技术的煤炭生产数据处理方法的执行主体可以为数据处理系统,该系统可以由软件和/或硬件的方式实现,该系统可以配置在电子设备中。示例性的,该电子设备可以包括但不限于终端、服务器端等。

[0033] 图1为本申请实施例所提供的基于物联网技术的煤炭生产数据处理方法的流程示意图。如图1所示,该基于物联网技术的煤炭生产数据处理方法可以包括但不限于如下步骤。

[0034] 在步骤101中,基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据。

[0035] 在一些实施例中,边缘网关可以从煤炭行业生产现场的PLC(Programmable Logic Controller,可编程逻辑控制器)、DCS(Distributed Control System,分散控制系统)、SCADA(Supervisory Control And Data Acquisition,数据采集与监视控制系统)等设备或系统中采集产线测点数据作为该生产数据。示例性的,该边缘网关中可以包括采集模块,通过该采集模块从边缘网关接入设备采集煤炭行业的生产数据。

[0036] 在一些实施例中,边缘网关接入设备可以包括但不限于以下一个或多个设备:安全监测分站、水文监测分站、压力监测分站、应力监测仪、色谱分析仪、拾震器、水泵等。可选的,该边缘网关接入设备还可包括接入传感器,示例性的,该接入传感器可以包括:氧气浓度传感器、一氧化碳浓度传感器、瓦斯浓度传感器、硫化氢浓度传感器、二氧化碳浓度传感器、粉尘浓度传感器、水温传感器、水压传感器、水位传感器、流量传感器、氮气浓度、乙烯浓度、乙烷浓度、风速、风压、温度、馈电状态、顶板压力、围岩应力、水泵流量、水泵功率等各类传感器和数据点。

[0037] 在一些实施例中,采集的煤炭行业的生产数据可以包括时序数据,可选的,该生产数据还可以包括非时序数据,可以根据应用系统的实际需求来决定采集的生产数据是时序数据还是非时序数据,或者既包括时序数据又包括非时序数据。可选的,当边缘侧基础网络不能保障数据实时传输时,可选择在数据处理系统中的边缘侧部署边缘时序库,用于缓存生产产线产生的大量时序数据,边缘时序库一般不会存储太长周期的时序数据,示例性的,可以存储一段时间(如5至7天)内的时序数据,也可以根据生产实际情况调整增加配置来达到更长周期数据的存储。边缘时序库中的数据可以根据网络情况定期将数据压缩并传递至数据处理平台侧。

[0038] 在步骤102中,获取与生产数据关联的特征信息,并获取数据处理系统的响应信息。

[0039] 在本申请的实施例中,可以在基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据的过程中,记录生产数据的关键信息要素,基于该关键信息要素确定与生产数据关联的特征信息。在一些实施例中,与生产数据关联的特征信息可以包括但不限于以下至

少一项:单位生产数据的处理时长、处理的单位生产数据数量、处理的生产数据总量、生产数据系统处理的复杂度、处理的生产数据最大延迟时间。

[0040] 在本申请的实施例中,可以从数据处理系统中读取该数据处理系统的响应信息。在一些实施例中,上述数据处理系统的响应信息可以包括但不限于以下至少一项:系统处理数据的基准时长、系统处理单位生产数据的基准时长、系统实际延迟。

[0041] 在步骤103中,根据特征信息和响应信息,确定数据处理系统处理生产数据的处理效率。

[0042] 在一些实施例中,可以根据生产数据关联的特征信息和数据处理系统的响应信息,基于深度学习技术预测数据处理系统处理生产数据的处理效率。示例性的,可以利用训练数据对深度学习模型进行训练,从而得到训练好的深度学习模型,该训练数据可以包括历史生产数据关联的特征信息和数据系统的响应信息,其中,该深度学习模型的输入包括历史生产数据关联的特征信息和数据系统的响应信息,输出为数据处理系统处理生产数据的处理效率;该深度学习模型可以为卷积神经网络模型,但不限于此。在模型应用中,可以将获得的生产数据关联的特征信息和数据处理系统的响应信息,输入至预先训练好的深度学习模型中,从而可以得到数据处理系统处理生产数据的处理效率。

[0043] 在一些实施例中,可以根据生产数据关联的特征信息和数据处理系统的响应信息,确定多个实时处理效能信息;根据多个实时处理效能信息,确定数据处理系统处理生产数据的处理效率。其中,在一些实施例中,上述多个实时处理效能信息可以包括但不限于以下至少一项:实时处理率、实时接收率、实时处理数据量、实时处理复杂度、实时延迟率。

[0044] 在一种可能的实现方式中,可以根据单位生产数据的处理时长和系统处理数据的基准时长,确定上述实时处理率。示例性的,可以将单位生产数据的处理时长与系统处理数据的基准时长的比值进行对数运算,得到的值确定为实时处理率。例如,该实时处理率的计算公式可如下表示: $r1 = \ln(t1/t3)$,其中, $r1$ 为实时处理率, $t1$ 为单位生产数据的处理时长, $t3$ 为系统处理数据的基准时长, $\ln()$ 为以常数 e 为底数的对数函数。当 $t1 = t3$ 时, $r1$ 值为0,其可以作为阈值点,单位生产数据的处理时长 $t1$ 越长,则 $r1$ 值越大,该生产数据实时处理需求越高。

[0045] 在一种可能的实现方式中,可以根据上述处理的单位生产数据数量和系统处理单位生产数据的基准时长,确定上述实时接收率。示例性的,可以将上述处理的单位生产数据数量与系统处理单位生产数据的基准时长的比值,确定为上述实时接收率。例如,该实时接收率的计算公式可以表示如下: $r2 = n/t4$,其中, $r2$ 为实时接收率, n 为处理的单位生产数据数量, $t4$ 为系统处理单位生产数据的基准时长。

[0046] 在一种可能的实现方式中,可以根据上述处理的生产数据总量和系统处理数据的基准时长,确定上述实时处理数据量。示例性的,可以将该处理的生产数据总量与系统处理数据的基准时长的比值,确定为该实时处理数据量。例如,该实时处理数据量的计算公式可以表示如下: $v1 = v/t3$,其中, $v1$ 为实时处理数据量, v 为该处理的生产数据总量, $t3$ 为系统处理数据的基准时长。

[0047] 在一种可能的实现方式中,可以根据生产数据系统处理的复杂度和系统处理单位生产数据的基准时长,确定上述实时处理复杂度。示例性的,可以将该生产数据系统处理的复杂度与系统处理单位生产数据的基准时长的比值,确定为该实时处理复杂度。例如,该实

时处理复杂度的计算公式可以表示如下： $c_1 = c/t_4$ ，其中， c_1 为实时处理复杂度， c 为生产数据系统处理的复杂度， t_4 为系统处理单位生产数据的基准时长。

[0048] 在一种可能的实现方式中，可以根据处理的生产数据最大延迟时间和系统实际延迟，确定上述实时延迟率。示例性的，可以将该处理的生产数据最大延迟时间与系统实际延迟的比值进行对数运算，将得到的数值确定为该实时延迟率。例如，该实时延迟率的计算公式可以表示如下： $r_3 = \ln(t_2/t_5)$ ，其中， r_3 为实时延迟率， t_2 为处理的生产数据最大延迟时间， t_5 为系统实际延迟， $\ln()$ 为以常数 e 为底数的对数函数。

[0049] 在一种可能的实现中，在得到多个实时处理效能信息，可以基于该多个实时处理效能信息，采用深度学习技术预测数据处理系统处理生产数据的处理效率。示例性的，可以将历史数据(如一段时间内的处理效能信息)作为训练数据，基于该训练数据对深度学习模型进行训练，利用训练好的深度学习模型预测数据处理系统处理生产数据的处理效率。

[0050] 在另一种可能的实现方式中，可以根据多个实时处理效能信息，采用第一公式计算数据处理系统处理生产数据的处理效率；其中，该第一公式可以表示如下：

$$[0051] \quad E = \frac{p_1 * r_1 + p_2 * \arcsin(r_2) + p_3 * v_1 * c_1 + p_4 * r_3}{p_1 + p_2 + p_3 + p_4}$$

[0052] 其中， E 为数据处理系统处理生产数据的处理效率； r_1 为实时处理率， r_2 为实时接收率， v_1 为实时处理数据量， c_1 为实时处理复杂度， r_3 为实时延迟率，均为实时处理效能信息； p_1 、 p_2 、 p_3 、 p_4 均为处理效能影响因子。示例性的，上述 p_1 、 p_2 、 p_3 、 p_4 可以是预先设定的可调节参数，例如可以基于大量试验得到的经验值。或者，示例性的，上述 p_1 、 p_2 、 p_3 、 p_4 可以是超参数，经过深度学习技术而学习得到的。

[0053] 在步骤104中，从多种处理方式中选用与处理效率匹配的处理方式，处理生产数据。其中，多种处理方式包括流式数据处理方式、批量数据处理方式和混合处理方式，混合处理方式是指混合流式数据处理和批量数据处理的方式。

[0054] 在本申请的实施例中，可以将处理效率与第一阈值和第二阈值进行大小比对，基于大小比对的结果，从多种处理方式中选用对应的处理方式对生产数据进行处理。

[0055] 在一些实施例中，在处理效率小于第一阈值的情况下，可以采用与处理效率匹配的批量数据处理方式处理生产数据。示例性的，可以利用批量计算工具对生产数据进行数据加工与处理，并将处理完成的数据存储至数据仓库中。例如，当数据处理系统处理生产数据的处理效率小于第一阈值时，可以认为该采集的生产数据是非时序数据(不是时序数据)或其他原因导致该生产数据需要批量处理，此时可以使用Spark批量计算工具对生产数据进行加工与处理，处理完成的数据直接存储到数据仓库进行永久保存。示例性的，数据仓库数据可以适用于需要存储大量历史数据并进行深入分析的业务系统。

[0056] 在一些实施例中，在处理效率大于第二阈值的情况下，可以采用与处理效率匹配的流式数据处理方式处理生产数据。示例性的，将生产数据存储至Kafka集群中，并基于实时计算工具对Kafka集群中的生产数据进行数据运算与处理，将处理完成的数据存储至时序数据库；其中，时序数据库中的数据定期汇总至数据仓库中进行保存。例如，当数据处理系统处理生产数据的处理效率大于第二阈值时，可以认为该采集的生产数据是时序数据，需要高效计算，如需要采用流式数据处理方式进行处理，比如可以使用Kafka集群接收流式数据(即采集的生产数据)，利用Flink等实时计算工具对Kafka集群中的生产数据进行数据

运算与处理,处理完成的数据将存储于时序数据库中,并定期汇总到数据仓库进行长期存储。示例性的,时序数据库数据可以适用于时效性要求高但不需要大量历史数据的业务系统。

[0057] 在一些实施例中,在处理效率大于或等于第一阈值且小于或等于第二阈值的情况下,可以采用与处理效率匹配的混合处理方式处理生产数据。在一种可能的实现方式中,可以确定混合处理方式的数据分配率;基于该数据分配率,将生产数据拆分成第一部分生产数据和第二部分生产数据,其中,第一部分生产数据待用流式数据处理方式处理,第二部分生产数据待用批量数据处理方式处理。可以根据不同业务需求,时序数据库和数据仓库可以灵活结合,以满足复杂的应用场景。

[0058] 在一种可选的实现方式中,上述数据分配率可以是预先设定的固定值。在另一种可选的实现方式中,该数据分配率可以是基于实时处理数据量和实时处理复杂度确定的,示例性的,可以维护一个数据处理对应表,该表中包括 $v1*c1$ 与时序数仓使用比值之间的映射关系,示例性的,该时序数仓使用比值可以为 $[0, 1]$ 的值(即该时序数仓使用比值的取值范围可以为大于或等于0且小于或等于1),该 $v1*c1$ 表示实时处理数据量与实时处理复杂度的乘积值,例如,假设从该表中找出 $v1*c1$ 对应的时序数仓使用比值为0.25,则确定混合处理方式的数据分配率为1:4,即混合处理该生产数据流式数据处理和批量数据处理的分配率为1:4,可以将该生产数据中的五分之一数据作为上述第一部分生产数据进行流式数据处理方式处理,将该生产数据中的五分之四数据作为上述第二部分生产数据进行批量数据处理方式处理。

[0059] 在一些实施例中,在采用与处理效率匹配的处理方式处理生产数据的过程中,可以基于预设的预警规则判断是否触发报警事件,并在触发报警事件时,基于报警事件关联的报警方式进行报警。示例性的,在对生产数据进行流式数据处理和/或批量数据处理的过程中,可以基于预警规则进一步判断是否会触发报警事件,如果触发报警事件,则可以基于该报警事件关联的报警方式进行报警。其中,该预警规则可以根据实际需求进行设置,例如该预警规则中可以包括报警事件的触发条件,还可以包括报警事件关联的报警方式等,本申请对此并不作具体限定,也不再赘述。

[0060] 在上述实施例中,可以根据采集的煤炭行业生产数据关联的特征信息和数据处理系统的响应数据,确定数据处理系统处理该生产数据的处理效率,便于选用与该处理效率匹配的处理方式来处理该生产数据,使得煤炭生产数据能够高效且精准地进行采集和计算,为煤炭行业的智能化生产、质量监控、环境监测等提供坚实的基础。

[0061] 图2为本申请实施例提供的数据处理系统的框图。如图2所示,该数据处理系统可以包括:采集模块201和数据处理平台202。其中,采集模块201,用于基于边缘网关从边缘网关接入设备采集煤炭行业的生产数据,并获取与生产数据关联的特征信息和数据处理系统的响应信息,根据特征信息和响应信息,确定数据处理系统处理生产数据的处理效率。数据处理平台202,用于从多种处理方式中选用与处理效率匹配的处理方式,处理生产数据;其中,多种处理方式包括流式数据处理方式、批量数据处理方式和混合处理方式,混合处理方式是指混合流式数据处理和批量数据处理的方式。示例性的,该采集模块201可以部署在边缘网关中。

[0062] 在一些实施例中,采集模块201用于:根据特征信息和响应信息,确定多个实时处

理效能信息;根据多个实时处理效能信息,确定数据处理系统处理生产数据的处理效率。示例性的,特征信息包括以下至少一项:单位生产数据的处理时长、处理的单位生产数据数量、处理的生产数据总量、生产数据系统处理的复杂度、处理的生产数据最大延迟时间;响应信息包括以下至少一项:系统处理数据的基准时长、系统处理单位生产数据的基准时长、系统实际延迟;多个实时处理效能信息包括以下至少一项:实时处理率、实时接收率、实时处理数据量、实时处理复杂度、实时延迟率。

[0063] 在一些实施例中,采集模块201用于:根据单位生产数据的处理时长和系统处理数据的基准时长,确定实时处理率;根据处理的单位生产数据数量和系统处理单位生产数据的基准时长,确定实时接收率;根据处理的生产数据总量和系统处理数据的基准时长,确定实时处理数据量;根据生产数据系统处理的复杂度和系统处理单位生产数据的基准时长,确定实时处理复杂度;根据处理的生产数据最大延迟时间和系统实际延迟,确定实时延迟率。

[0064] 在一些实施例中,采集模块201用于:根据多个实时处理效能信息,采用第一公式计算数据处理系统处理生产数据的处理效率;其中,第一公式表示如下:

$$[0065] \quad E = \frac{p1 * r1 + p2 * \arcsin(r2) + p3 * v1 * c1 + p4 * r3}{p1 + p2 + p3 + p4}$$

[0066] 其中,E为数据处理系统处理生产数据的处理效率;r1为实时处理率,r2为实时接收率,v1为实时处理数据量,c1为实时处理复杂度,r3为实时延迟率,均为实时处理效能信息;p1、p2、p3、p4均为处理效能影响因子。

[0067] 在一些实施例中,数据处理平台202用于:处理效率小于第一阈值,采用与处理效率匹配的批量数据处理方式处理生产数据;或者,处理效率大于第二阈值,采用与处理效率匹配的流式数据处理方式处理生产数据;或者,处理效率大于或等于第一阈值且小于或等于第二阈值,采用与处理效率匹配的混合处理方式处理生产数据。

[0068] 在一些实施例中,数据处理平台202用于:确定混合处理方式的数据分配率;基于数据分配率,将生产数据拆分成第一部分生产数据和第二部分生产数据,其中,第一部分生产数据待用流式数据处理方式处理,第二部分生产数据待用批量数据处理方式处理。

[0069] 在一些实施例中,数据处理平台202用于:利用批量计算工具对生产数据进行数据加工与处理,并将处理完成的数据存储至数据仓库中。在一些实施例中,数据处理平台202用于:将生产数据存储于Kafka集群中,并基于实时计算工具对Kafka集群中的生产数据进行数据运算与处理,将处理完成的数据存储至时序数据库;其中,时序数据库中的数据定期汇总至数据仓库中进行保存。

[0070] 在一些实施例中,数据处理平台202还用于:在采用与处理效率匹配的处理方式处理生产数据的过程中,基于预设的预警规则判断是否触发报警事件,并在触发报警事件时,基于报警事件关联的报警方式进行报警。

[0071] 需要说明的是,前述对基于物联网技术的煤炭生产数据处理方法实施例的解释说明也适用于该实施例的数据处理系统,此处不再赘述。

[0072] 为了方便本领域技术人员更加清楚地了解本申请,下面将结合图3进行描述。

[0073] 如图3所示,该数据处理系统可以分为边缘侧、平台侧和应用侧。其中,边缘侧可以部署边缘网关,边缘网关从生产现场的PLC、DCS、SCADA等设备或系统中采集产线测点数据,

采集数据可以通过4G/5G移动网络或者企业内部光纤网络发送至采集平台侧(如数据处理平台)。例如,可以从业务系统(如办公自动化OA系统、工程生产管理系统PMS、网络内容控制系统NCC系统、知识库系统、人力系统等)中采集业务系统数据作为该生产数据。

[0074] 当边缘侧基础网络不能保障数据实时传输时,可选择在边缘侧部署边缘时序库,用于缓存生产产线产生的大量时序数据,边缘时序库一般不会存储太长周期的时序数据,通常存储5至7天内的时序数据,也可以根据生产实际情况调整增加配置来达到更长周期数据的存储。边缘时序库中的数据会根据网络情况定期将数据压缩并传递至平台侧。

[0075] 示例性的,边缘网关接入设备包括:安全监测分站、水文监测分站、压力监测分站、应力监测仪、色谱分析仪、拾震器、水泵等各类设备。涉及的接入传感器包括:氧气浓度传感器、一氧化碳浓度传感器、瓦斯浓度传感器、硫化氢浓度传感器、二氧化碳浓度传感器、二氧化碳浓度传感器、粉尘浓度传感器、水温传感器、水压传感器、水位传感器、流量传感器、氮气浓度、乙烯浓度、乙烷浓度、风速、风压、温度、馈电状态、顶板压力、围岩应力、水泵流量、水泵功率等各类传感器和数据点。

[0076] 如图3所示,平台侧可以对采集的生产数据(如时序数据、非时序数据)进行计算与存储。基于边缘网关采集的生产数据,可以根据实际应用场景划分为两种情况,对于流式数据处理的情况,数据首先会存储在Kafka集群中,利用Flink实时计算工具进行数据运算与处理,处理完成的数据会存储在时序数据库(如InfluxDB或CirroTimes)中,时序库中的数据会定期汇总至数据仓库中进行永久保存。对于批量数据处理的情况,利用Spark批量计算工具进行数据加工与处理,处理完成的数据会直接存储在数据仓库(如达梦数据库或CirroData)中,进行永久保存。

[0077] 通常生产数据采集到平台侧后,会有实时预警的需求,实现对生产中,Flink和Spark会读取关系库(如MySQL)中的预警规则,来进一步判断是否需要报警以及报警的方式等。

[0078] 针对以上煤炭生产数据,主要采用以下三种处理方式:

[0079] (1) 流式数据处理:使用Kafka集群接收流式数据,利用Flink进行实时数据计算与处理。处理完成的数据将存储于时序数据库(如InfluxDB、CirroTimes)中,并定期汇总至数据仓库(如达梦数据库、CirroData)进行长期存储。

[0080] (2) 批量数据处理:使用Spark批量计算工具对数据进行加工与处理,处理完成的数据直接存储至数据仓库进行永久保存。

[0081] (3) 混合式处理:两种处理方式混合使用。

[0082] 对应的,生产数据的采集结果最终会存储在时序数据库和数据仓库中,供应用系统根据需求进行调用。其中,时序库数据:适用于时效性要求高但不需要大量历史数据的业务系统。数据仓库数据:适用于需要存储大量历史数据并进行深入分析的业务系统。混合使用:可以根据不同业务需求,时序库和数据仓库可以灵活结合,以满足复杂的应用场景。示例性的,可以通过本申请实施例涉及的方法实施例来对煤炭生产数据进行处理,其实现方式可参见上述方法实施例的可选实现方式,此处不再赘述。

[0083] 如图3所示,应用侧可以进行煤炭生产数据的使用。生产数据的采集结果会存储在时序数据库和数据仓库中,应用侧根据实际使用需求,选择获取哪个数据库中的数据即可。通常时序库中的数据会提供给时效性要求较高,但不是很关心海量历史数据的业务系统。

而对于时效性要求不高,但对海量历史需要有较高要求的业务系统,会选择获取数据仓库中的数据。根据实际业务需求,时序库与数据仓库也可以混合使用,从而满足更复杂的业务需求。

[0084] 值得注意的是,本申请实施例可以实现对生产数据(如生产环境时序数据)的采集、计算、存储提供有力支撑。并且这些生产数据未来的应用场景十分广阔,不仅对数据挖掘与深度应用提供支撑基础,同时对未来的智能工厂、自动化生产、生产工艺优化、产品质量监控和追溯都起着至关重要的作用,这些应用场景的实现将有助于提高生产效率、降低成本、提升产品质量和市场竞争力。

[0085] 为了实现上述实施例,本申请还提出一种电子设备,包括:至少一个处理器;以及与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行本申请所提供的基于物联网技术的煤炭生产数据处理方法。

[0086] 为了实现上述实施例,本申请还提出一种存储介质,存储介质存储有指令,当指令在电子设备上运行时,使得电子设备执行本申请所提供的基于物联网技术的煤炭生产数据处理方法。

[0087] 为了实现上述实施例,本申请还提出一种计算机程序产品,当计算机程序产品中的指令被电子设备中的处理器执行时,实现本申请所提供的基于物联网技术的煤炭生产数据处理方法的步骤。

[0088] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0089] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本申请的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本申请的实施例所属技术领域的技术人员所理解。

[0090] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤,例如,可以被认为用于实现逻辑功能的可执行指令的定序列列表,可以具体实现在任何计算机可读介质中,以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用,或结合这些指令执行系统、装置或设备而使用。就本说明书而言,“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下:具有一个或多个布线的电连接部(电子装置),便携式计算机盘盒(磁装置),随机存取存储器(RAM),只读存储器(ROM),可擦除可编程只读存储器(EPROM或闪速存储器),光纤装置,以及便携式光盘只读存

储器 (CDROM)。另外,计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质,因为可以例如通过对纸或其他介质进行光学扫描,接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序,然后将其存储在计算机存储器中。

[0091] 应当理解,本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。如,如果用硬件来实现和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或它们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列 (PGA),现场可编程门阵列 (FPGA) 等。

[0092] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0093] 此外,在本申请各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读取存储介质中。

[0094] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。尽管上面已经示出和描述了本申请的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本申请的限制,本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

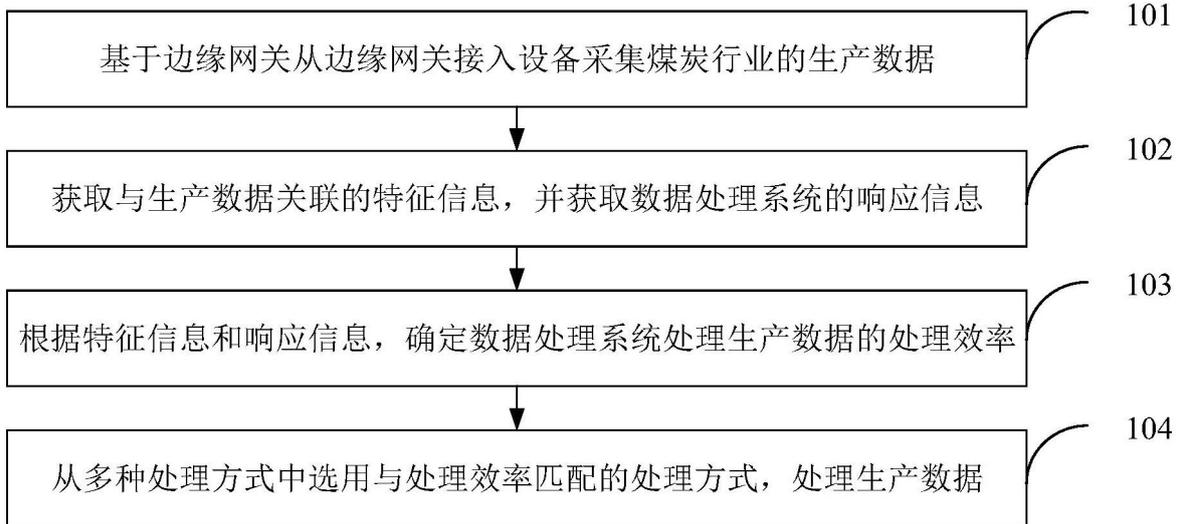


图1

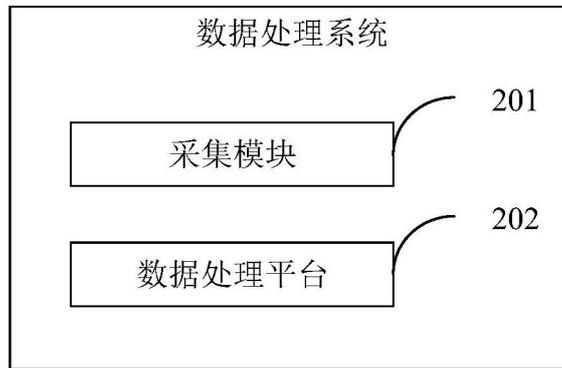


图2

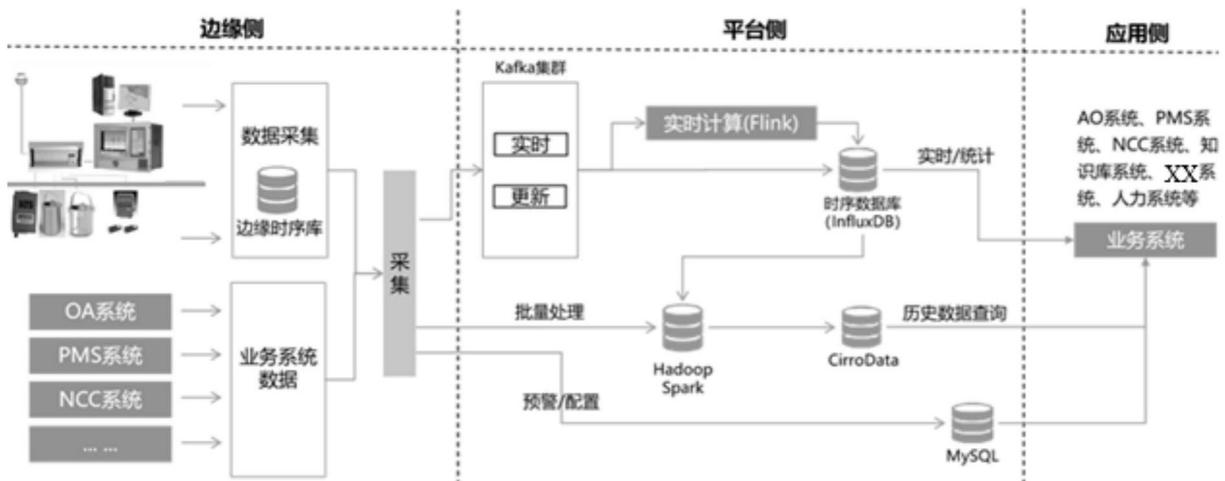


图3