

一种物联网数据管理框架研究

史俊茹 黑敏星 杨 军

(宁夏大学数学计算机学院 银川 750021)

摘 要 物联网是一种由智能实体持续不断地产生数据、物体和物体之间互联互通以及通过互联网将数据进行传输的网络范例。目前,管理和利用由这样一类实体产生的海量数据的解决方案尚不成熟。传统的数据管理解决方法在满足物联网复杂的应用需求方面存在缺陷和不足,主要是因为受到物联网的全球规模限制。针对物联网数据管理,当前关注的焦点主要是传感器网络。主要探究了关于物联网或物联网各子系统的数据管理解决方案。最终提出一种关于物联网的数据管理框架,它采用一种联邦的、以数据源为中心的方式连接拥有丰富数据的实体,并将其作为综合性物联网数据管理方式的一种探讨,为未来物联网的潜在应用和服务提供支撑。

关键词 物联网,数据管理,联邦,传感器网络

中图法分类号 TP393 **文献标识码** A

Data Management Framework for Internet of Things

SHI Jun-ru HEI Min-xing YANG Jun

(School of Mathematics and Computer Science, Ningxia University, Yinchuan 750021, China)

Abstract The Internet of things(IoT) is a networking paradigm where interconnected smart objects continuously generate data and transmit it over the Internet. However, the solutions to manage and utilize the massive volume of data produced by these objects are yet to mature. Traditional database management solutions fall short in satisfying the sophisticated application needs of an IoT network that has a truly global-scale. Current solutions for IoT data management address partial aspects of the IoT environment with special focus on sensor networks. We surveyed the data management solutions proposed for IoT or subsystems of the IoT. And finally proposed a data management framework for IoT that acts as a seed to a comprehensive IoT data management solution. The proposed framework adapts a federated, data-and sources-centric approach to link the diverse things with their abundance of data to the potential applications and services that are envisioned for IoT.

Keywords Internet of things, Data management, Federation, Sensor networks

1 引言

物联网是一种动态、全球化的网络基础结构。其中,物联网实体可由子系统、单个或虚拟物体等构成,这些实体皆具有可标识、自动化以及可自我配置的特性。实体间能够相互通信,即实体与环境间通过交换感知的数据进行交互,并对监测事件作出响应,产生与之相适应的控制物理世界措施。物联网试图实现的是提供一种标准化的平台,用来开发合作性的服务和应用程序,这些服务和应用程序通过单个实体和任何子系统(管理先前提到的实体),对适用资源进行挖掘和分析。这些适用资源蕴含丰富的信息,而这些信息又可通过数据融合等技术变得更加实用,从而使得创新和非常规应用以及增值服务有望实现。同时,也提供一种宝贵的趋势分析和战略机遇。因此,针对物联网环境下实体对象生成和存储的海量数据,需要一个综合型的数据管理框架实现以上目标。

数据管理是一个广泛的概念,涉及体系结构、关于特定系

统数据生命周期的恰当管理流程等。在物联网的感知上下文中,数据管理应该扮演的是一个层面,它位于生成数据的设备对象与获取数据用于分析的应用程序之间。设备间可通过自动化自我治理或是内部层次管理,自组织成各个子系统或是子空间,然后这些子系统所提供的数据和功能服务于物联网。

物联网数据有其独到的特性,这些特性使得传统的基于关系的数据管理成为一种不适应、过时的解决方案。成千上万多种多样的设备产生海量异构、流模式以及地理位置分散的实时数据,且这些设备周期性地发送关于特定监控现象的观察结果,亦或是报告感兴趣的特定/异常事件的发生情况。因此,在设计物联网数据管理解决方案时,通信、存储和处理将成为主导性因素。

2 相关工作

当前被提出的大部分数据管理主要是针对全球化物联网领域的子集——无线传感器网络(WSNs),而这些并不能很

本文受国家自然科学基金项目(61261001),教育部科学技术研究重点项目(212189)资助。

史俊茹(1989—),女,硕士生,主要研究领域为无线传感器网络,E-mail:shijunruws@sina.cn;黑敏星(1989—),男,硕士生,主要研究领域为无线传感器网络;杨 军(1972—),男,博士,硕士生导师,CCF 会员,主要研究领域为无线传感器网络,E-mail:dragon@nxu.edu.cn(通信作者)。

确切地适用于物联网更加复杂的结构特征。无线传感器网络是发展成熟的网络典范,其数据管理方案主要是围绕内网的数据处理和优化。传感器大部分都具有静态、资源受限的特性,而这些特性并不有利于复杂的分析和数据。基于 WSNs 的数据管理方式,其主要的重点是利用有限的固定存储能力实时迅速地获取数据以利于快速决策,达到长期使用^[1]。而这仅仅是具有更全面功能的物联网系统的子集,物联网旨在综合处理和利用从各种资源获取的可适用数据,这些资源包括:静态和动态的、智能与嵌入式的、资源受限和资源丰富的、实时以及归档在案的等。因此,为了寻求有益的全球模式和战略机遇,基于物联网数据管理方式,其重点是拓展和延伸无线传感器网络所能提供的各种功能和服务,并且能够通过挖掘海量的异构数据以一种无缝接入的方式增加更多的服务功能。

文献[2,3]探讨了“物联网的标志性特征”和“物联网的模式描述”两个基本问题;指出物联网的主要特征在于:自反馈体系架构、3C、安全、复杂网络、复合生态系统,并给出以时、空、量、构、序、信为主体的物联网复合生态系统理论模型。

赵志军^[4]等着重分析研究了智能信息处理的信息空间定义、信息量化方法以及信息处理各阶段的主要问题及相应的解决办法。童恩栋等^[5]主要阐述了物联网环境下情景感知建模、应用及系统结构等方面的问题,明确了情景感知的主要流程并介绍了各流程相关技术。并对现有的系统结构进行了分析和对比,结合物联网环境论述了当前情景感知系统的不足之处,并给出了情景感知系统的参考结构。

文献[6]针对物联网环境下云数据安全性问题,在云计算中为了保证用户数据的准确性和隐私性,提出了一种物联网环境下云数据存储安全及隐私保护策略。相对于传统的基于文件复制的分配技术,此策略降低了需要交换的信息量和数据存储量。利用同型数据分布式检测令牌,保证了数据的准确性并实现了错误定位。

文献[7]描述数据管理与智能处理是物联网必须解决的关键问题,鉴于物联网数据的多源、异构、海量、动态等特点以及物联网的可扩展性,文中提出了一种综合运用先进的数据空间技术、不确定数据管理技术、云计算技术解决这一问题的思路,也给出了在此思路下需要研究的关键问题。

文献[8]针对物联网中数据的特点、组成和数据的流转过,从数据模型、数据预处理与集成、数据查询、存储索引等几个方面,对物联网数据管理的相关技术进行了研究,提出了物联网环境下基于云计算平台的数据管理框架。框架主要包括数据预处理与集成、分布式的数据中心,其中数据中心要解决数据模型、数据存储与索引、查询分析处理等问题。

丁志明等^[9]针对物联网传感器采样数据管理中所面临的数据海量性、异构性、时空敏感性、动态流式特性等问题,提出一种面向物联网海量传感器采样数据管理的数据库集群系统框架 IoT-ClusterDB。为物联网海量异构传感器采样数据的存储与查询处理提供了一种可行的解决方案。

当前的一些建议侧重点都是提出抽象化来支持异构网络数据的集成,从而为其他物联网子系统的自适应与无缝集成铺平道路。然而,支持不同子系统间的互操作性,能够对集成移动对象、满足内容感知需求的数据整体生命周期进行管理,

综合而又全面的数据管理解决方案尚未得到开发。

本文提出一种框架用于物联网数据管理,该框架更符合物联网数据生命周期的过程。提出的框架采用分层的方式,以数据-源中心(data-and sources-centric)中间件为核心,并借用了联邦数据库管理系统的概念来保证相互不受约束的物联网子空间的独立性和子系统加入/离开架构体系的灵活性。

3 物联网数据管理框架描述

物联网典型网络分层结构为 5 层:感知层、接入层、网络层、支撑层和应用层。本文提出的物联网数据管理框架由 6 个堆叠层组成,其中有两个堆叠层包含子层或互补的双层。框架的各个层间紧密地映射到数据的整个生命周期中,如图 1 所示。感知识别层包含物联网传感器和智能对象(数据生成对象),以及内网处理模块和数据采集/实时聚合(加工、聚合)模块。通信传输层支持各种请求、查询、数据和结果(收集和交付)的通信及传输。资源接入层处理数据来源的发现和编目,并对收集后的数据(数据存储/档案)进行存储和建立索引。数据层同样也进行数据的查询处理,主要是针对本地自动数据存储库站点的数据进行过滤、预处理和处理。数据联邦层提供了必要的数据存储库的集成和抽象化,这样处理的目的是为了使用存储在资源接入层的元数据、支持实时的资源集成以及处理全球化查询/分析请求(预处理、集成、融合)。

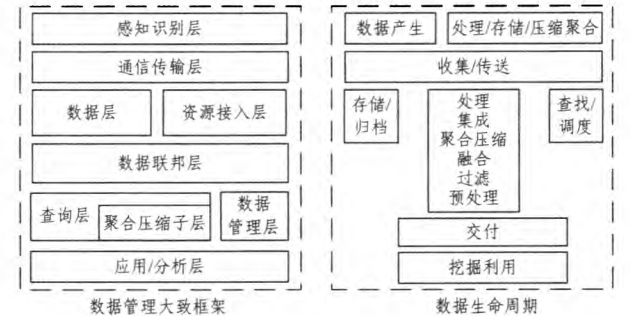


图 1 数据管理框架各个堆叠层与数据生命周期的相互映射^[1]

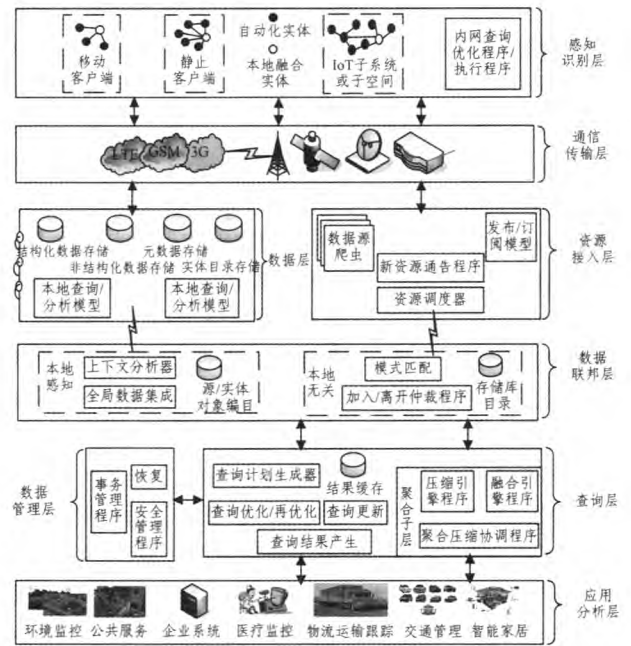


图 2 数据管理框架体系结构

¹⁾ 图 1 中的查找/调度被认为是一个额外的过程,严格来讲并不是数据生命周期的一部分。

查询层主要是负责查询处理和优化,并与联邦层相互合作、与事务管理层进行互补操作(加工、交付)。查询层包含聚合子层,用来处理涉及一系列数据源/站点的数据聚合或融合查询。应用分析层是数据/分析需求的请求者,且是数据和分析结果的消费者。框架中各个层的功能模块如图2所示,并将在接下来详细描述。

3.1 数据层

由于数据层是数据管理的核心元素,因此先对该模块功能进行阐述。理解数据存储在哪里以及如何存储,对后续数据的更新、查询和访问来说,具有至关重要的作用。物联网数据管理就数据本身而言,有两个主要问题需要解决:存储设施的位置(Where)和用于数据存储的格式(How)^[1]。在提出的框架中,选择混合的数据存储方式:与时序有关的实时数据,将其存储在生成数据实体的附近站点;而用于记录分析持久、归档的长期数据则储存在专用的设施上。以期通过这样的形式,在存储空间、数据传输成本与数据(针对复杂的分析、查询)的可用性之间获得折中的收益效果。数据层与持久稳定的数据存储相关联,而感知识别层关心的主要是临时数据的存储,关于其他方面将在后续的小节中讨论。

由于应用程序将通过访问物联网系统来充分利用数据,因此数据流向的设计元素需要将数据属性和应用程序的需求考虑在内。针对物联网全球化、生成数据的实体位置各异(有些在地理位置上可能相隔甚远)等特性,故适宜采用联邦方法对数据进行存储和实时访问。联邦数据管理包含本地数据源的自主控制,当本地数据源上的资源允许被访问时,本地数据源可以选择性地参与联邦系统执行的全局查询。联邦系统执行全局查询可能涉及到各种数据源,但数据源处理数据请求的方式并不受此影响。同时,当数据源进入或离开联邦体时,亦不会影响本地系统的一致性^[10]。类似的方法可以应用到物联网数据管理中,数据由不同的子系统生成,或由一组被子系统的所有者放置在指定位置的相关对象生成。当某一本地子系统需要查询访问相关联的其他地理位置遥远的子系统时,不同的子系统可以参与到联邦数据库中,前提是它们能量受限的资源能够满足。因此,数据层可以被视为驻留在不同物联网存储站点的数据联盟的一种抽象表示,一方面本地模块来处理这些数据,另一方面目录来识别数据的规范,以方便在联邦层的数据联合。数据层的本地模块功能,与上层的数据处理功能类似。从本地同一个物联网存储站点层次上来看,本地查询/分析模块与查询层的查询处理和优化模块具有相同的执行功能,这些在后面的查询层进行探讨。本地数据集成模块对同一个自治物联网系统生成的数据执行简单的集成处理,这些数据可能是结构化或非结构化的,故而有必要对数据进行统一,以便于查询和结果分析。

关于数据层的实体目录存储,它的作用主要是获取数据和定义信息来源(生成信息的对象或设备),以利于信息的存储。为了高效地对数据进行复杂逻辑条件查询,且考虑结果的特定性和异构数据的集成,选择采用语义元数据来描述数据是如何被访问和被链接。数据可以通过空间地理位置和时

间标记序列关联起来,以促进语义的相关性以及建立位置和时间关联的上下文;其中,某些数据只有在特定的地点和特定的时间间隔内才是相互关联的。在实体目录中,定义数据源需采取一种分层的形式,类似于语义结构^[2]的来源,使用内联表示存在的依赖关系。另外,涉及到数据的精确程度方面,生成数据的各种源可与置信度关联起来,以利于查询优化。目录信息是用于提供异构数据存储间互操作模式映射的主要索引来源。特定域的描述也应是元数据不可缺少的一部分。

3.2 感知识别层

感知识别层包含各式各样的实体,这些实体生成的数据用于物联网系统以及内网执行模块进行实时处理,并将处理后的结果进一步交付到物联网系统中。实体可以是传感器、射频识别标签、移动设备、笔记本电脑、嵌入式芯片、车辆、船舶,以及任何装配有通信能力的嵌入式智能设备。智能设备或可称之为虚拟体的物组成的集合也可被看作是实体,如病人身上佩戴有小型区域网络、装备有大量智能设备的车辆,或者一个环境传感器网络。这些实体对象,无论其能否自动化操作亦或是仅作为一个内联网络系统的一部分,都可被认为是物联网网络的数据源。

为了能够访问感知识别层的数据,需要有一种机制来唯一地标识数据源。因全球物联网的特性,针对在地理位置上分布式的物体所生成的数据,对其进行高效的检索和查询时,基于位置的身份识别是至关重要的。此外,模态识别也需要含并在感知识别层内,用来识别支持相同数据类型的同类实体。这类似于部分传感器网络所遵循的以数据为中心的命名机制。与平台相关联的身份识别可以与基于位置的身份识别相互参照。与平台相关联的身份识别用于识别属于特定平台且与实体或平台有关并用于服务通用目的的实体,正如身体区域网络与特定的病人相适应一样。因此,这种独特的识别是紧密依附于物联网对象,并能够通过上层框架的内网处理或者移动代理迅速灵敏地接入感知识别层。感知识别层有两个模块:本地聚合模块和网络查询优化模块。虽然这两个模块的复杂程度低,但其所起的作用与上层其他的模块类似。

本地聚合模块:优化传输与存储成本。由于聚合函数旨在减少通信成本,因此聚合点部署在靠近数据源时效果会更好。聚合模块将从多个对象和物体中接收的数据进行整理和摘取。聚合点的使用是为了方便处理和满足底层子系统/实体的高效率需求,并且不应导致时延,否则会降低系统的实时性能。

网络查询优化模块:主要是内层网络查询优化/执行模块。这个模块只有实时查询请求到达时才运行,并且只提供上层所需要的查询结果。这样的查询结果或以定期报告的形式永久存储在上层数据层中,或被及时执行以响应延迟敏感的查询请求。

3.3 通信传输层

通信传输层将分布式的数据源与更多的集中数据存储处理单元相连接。这需要用到对象-对象、对象-基础设施间的通信技术。同时,也需要为上层互操作性提供保障。当地理

^[1] 文献《Data Management for the Internet of Things: Design Primitives and Solution》的第18页有描述。

^[2] 语义结构含有施事、受事、谓词等成分,其最小单位是语义词(又叫义位),最大单位是义句。(来源于互动百科)

位置上分散的多个数据存储库被用于进行复杂的查询和结果分析时,通信传输也存在联邦层中。

3.4 资源接入层

在分布式数据库系统中,数据库碎片存储在预定义、有限的空间位置上。系统设计需事先指示出数据库碎片中元数据的存储位置,以用于之后的查询和更新。而对于物联网数据存储,这方面的情况差异很大。生成和传送数据的来源主要有以下几个特性:

- (1)分布式的,地理/平台位置各异。这些位置各式各样,可以是专用的对象,也可以是植入的对象;
- (2)不是有限的。有持续不断的可用新来源;
- (3)独立自主的,没有统一的模式,暂无统一的元数据定义。

鉴于数据来源的以上这些特性,如果没有办法识别哪些来源或子系统可以响应来自上层的查询请求,那么感知识别层执行实时查询就会变得很具有挑战性。因此需要位于感知识别层之上的一层,针对以查询处理为目的的请求,实现处理数据来源无缝且透明的识别和统一,即源层。另外,本层主要处理来自下层的请求,而不是集中于向上层周期性地汇报用于存储和归档的生成数据。

资源协调器:该资源协调器旨在解析一系列关于数据来源的查询执行指令。该模块同查询优化器紧密合作,用来查找与查询请求相匹配的源以及调整执行计划,以达到位置感知的效果。而且,本模块能建立合作方案以及遵循数据源查询结果交付时的层次结构。

数据源爬虫:物联网有着灵活的体系结构,因此能够容纳新的子系统无缝接入和激活运行。子系统合并入可扩展的物联网,可通过主动向物联网通告自己的存在;相反,也可在感知识别层通过运行发现例程被动地寻求新子系统的加入请求。动态活跃的新源可以被爬虫周期性地扫描发现,亦或只有当爬虫收到数据查询请求才进行扫描去发现新源。

新源通告程序:数据源爬虫发现新的数据源后,需要一种通告机制,通过该机制实时、持续不断地更新查询都能感知到新数据源的存在。另外,这些发现的新数据源,其数据规范需要被报告给数据层,主要是实现新数据源及时被编入数据层的实体目录中,并进行元数据的存储。新的子系统或者设备可直接将自己的规范说明报告给新源通告程序以被基础设施所知道。反过来,新源通告程序验证这些报告的真实性,并将通过验证的规范说明传送给地理位置最近或语义最相关的数据层实体目录存储模块和元数据存储模块。

发布/订阅模块:发布/订阅模块是在感知识别层与联邦层之间的中间代理。该模块接收数据的来源、数据的消费等信息,发布能为其他模块提供的数据/服务描述,以方便其他模块的查询和订阅。

3.5 数据联邦层

联邦层位于整个框架的中间部分,作为将分散的物联网子系统和各数据源连接起来的纽带,从而形成一个全球化的物联网系统。它提供了一种将不同数据类型和存储库联合起来响应某一特定查询的互操作。位置无关模块使得不同的源/子系统能够无缝地加入和离开联邦体;而位置感知模块处理明确要求具体地点位置的查询。

存储库/源目录:存储库目录持有不同的有望参与物联网数据存储库/数据库的元数据。源目录是存储库目录更加量化的一个版本,主要描述数据如何被定义。目录里包含元数据加入/离开网络的位置、时间以及数据的规范信息等。

加入/离开仲裁模块:在框架中,加入/离开仲裁程序支持联邦原则,用来管理数据库的纳入和排除。各存储库要求物联网基础设施的数据提供者宣布加入/离开仲裁模块,数据提供者需提供它们的数据模式、规范、所期望的性能和安全参数等。仲裁模块将存储库中相应的描述性信息添加到目录中。在任何时刻,通过定期公告或按需调查,仲裁模块了解哪些仓库正在积极用于参与查询/分析任务,并调度基于存储库的本地工作负载、处理优先级以及所需数据可用性等方面的特定查询和分析任务。

模式匹配模块:模式匹配模块应该提供一个统一连贯的、支持异构多样的、能被联邦层访问的数据存储库/数据库范型。同一块数据在不同的存储库拥有不同的名称时也是可被使用的,以及类似的实体在不同数据库模式中可有不同的表示形式。因此,模式匹配可以依靠比较属性值或定义数据属性的本体实现掩盖语义异构性和提供一个无缝的数据图解视图。事先定义好统一的全球匹配模式对动态数据存储库环境并不适用。因此,可以预定义基于模式信息模式级匹配的核心属性和对象类来支持数据级匹配。

上下文分析器:上下文分析器将提供上下文规范,支持来自上层少量对数据位置感兴趣的查询。上下文分析器能够判决最有可能拥有查询所需结果的数据源位置。上下文分析不仅仅能支持位置查询,同时也支持身份、状态、时间序列以及其他具有约束条件的相关查询。

数据集成模块:数据集成模块与查询层密切合作,通过处理包含在源目录中的信息,实现对异构数据的访问。数据集成将协助语义查询的执行,它还将不同的数据格式/语义映射到一个统一适用于查询结果的形式,提供给上层。无论数据是来自事先定义好的存储库还是来自实时生成数据的动态数据源,数据集成模块都统一提供用于查询的无缝结果格式。数据集成器部署在数据层,也为本地物联网子系统/站点提供相似的适应设备以便于处理本地查询。

3.6 查询层

查询层封装:用于物联网数据库生成、优化、执行查询所需的必要元素。该层部署在联邦和局部层面上。其中,局部层面指的是通过单个的组织或是代理在部署后管治各子系统。通过局部底层系统所生成的各个数据视图可以获得物联网数据的全局视图。

查询计划生成模块:查询计划生成模块需要获得详细的输入规范,才能实现理想的输出结果,这些详细的输入规范可来自应用程序或直接来自用户,并将它翻译成一种标准的查询语言形式。可能多个查询计划会以查询命令树或是查询步骤集的形式产生;查询的表现形式需要符合定义在数据库中的数据实体模式以及如何获取所需数据的详细指示说明。语义和上下文查询指示位置信息、时间间隔(序列)、特定业务语义需求也需要被解释到具体的可执行查询计划中。当然,这些会和数据适配器合作完成。针对某一查询的特定查询计划需被传送到查询优化器作进一步的处理,以期选择最优的查

询计划执行。

查询优化/再优化器:查询优化器接收关于某一特定查询的查询计划集合,对这些查询计划进行评估,最终找出在执行查询过程中最高效的一个。评估各个查询计划的优劣,可通过输入/输出的操作次数、查询处理的响应延迟以及基于时间序列、位置感知、数据源模式等约束条件下查询结果的数据准确度等指标进行比较。如果需要由用户发布查询请求,优化器可以针对动态的数据源直接执行“再优化”查询。

查询更新器:查询更新器是在各种查询到来时被激活执行,用于更新各种查询结果、周期性地更新生成的数据流、满足具有需要更新特性的服务以及及时通告哪些数据源动态加入/离开网络等。

聚合子层:该子层包含了3个模块,分别是聚合压缩协调程序、融合引擎程序以及压缩引擎程序。分布在全球不同地理位置的物联网数据存储库,时时刻刻都在更新海量数据,这些数据伴随时间向前推移。当复杂的查询需要访问实时生成的数据时,需要同联邦层合作以实现互操作,这就需要在查询层和联邦层之间有一个纽带层,而聚合子层便可实现此功能。聚合压缩协调程序主要负责从各种抽象的数据存储库系统中收集数据,并依据特定的排序对不同的存储库数据进行摘取。融合引擎程序负责将各种不同格式的数据进行融合,以提取数据摘要,同时融合函数包含在融合引擎程序内。压缩引擎程序实现的功能类似于感知识别层的本地聚合模块,主要是基于查询要求(时间、地点、模式上下文等等)对数据库中的数据进行摘取操作,以减少传输数据量。

3.7 管理层

管理层的主要功能是为存储在框架数据层中各式各样的数据提供所需的访问机制和安全机制。

事务管理器:事务处理器负责一些与业务进程或服务相关联的事务的具体执行。依据所提交上来的事务类型,事务管理器可以部署一个典型的单一执行机制,也可以部署全局式或分布式的多样执行策略。

恢复管理器:恢复管理器主要负责在出现断电、撞击事故或文件档案被损坏等情况后,将数据库恢复到一个最近且一致的状态。通常的做法是,回滚所有发生在数据管理系统中尚未提交的事务或操作。在物联网系统中,相同数据的适用性是来自多个相互独立的数据源,恢复可以通过潜在的“本源”复制实现。因此,恢复管理器只需关注独立的存储库维护,这样恢复仅需通过早已部署在数据库管理系统中的 Redo/Undo 恢复机制就能够得以实现。

安全管理器:安全管理器的主要功能是依照与受保护的数据和最终用户等相关的法律规范,对数据进行保护并实施隐私安全措施。物联网中,由于部分生成数据的物可能是私有实体或其他个人所有物,如生成数据的医疗设备或是车载系统等,为了使得信息不被非法窃取或是滥用,采取隐私安全保护措施是非常必要的,而且应给予高度重视。健壮的隐私安全保护措施与数据管理解决方案紧密相关联,也是物联网

成功部署的关键部分。

3.8 应用/分析层

应用/分析层主要关注由物联网所产生信息的利用情况,提供一系列服务和分析性能给物联网系统的终端用户。物联网中存在大量的数据存储部署策略,其主要目标是利用海量的信息进行相关的数据研究,挖掘潜在感兴趣的模式或发展趋势。对潜在价值的信息进行挖掘和发现可有助于系统正常运行、改进,并提供技术创新或非传统商业机遇。应用/分析主要涉及部署在联邦和本地层面的模块,以此能够提供全球物联网数据的不同视图。同时,促进本地自主访问或利用与物联网子系统相关的数据。目前,模式识别和数据挖掘等技术被用来开发众多的物联网应用。基于物联网数据开发各种具体的应用,数据的3个主要方面因素需被考虑:增加的大规模数据项;数据本身的非结构性和异构性以及基于地理位置的分布式存储设施。

结束语 本文从数据的采集、存储和处理3个方面出发,依据数据在整个生命周期中的状态,提出一种在物联网环境下综合型的数据管理框架。框架包含数据层和源层以及联邦体系结构等核心组件,并突出双向需求和跨层设计来满足实时/归档查询和应用分析/服务。此外,文中所提出的框架中,还有待需要考虑的地方,如在动态和异构的物联网环境中,数据的安全和隐私如何被保证;针对不同的数据类型和格式,能否构造有效的基于时间和地理位置的标签信息,实现物联网中异构数据源的集成等,都是今后的研究方向。

参 考 文 献

- [1] Abu-Elkheir M, Hayajneh M, Ali N A. Data Management for the Internet of Things: Design Primitives and Solution[J]. Sensors, 2013, 13: 15582-15612
- [2] 王瑞刚. 物联网主要特征与基础理论研究[J]. 计算机科学, 2012, 39(6A): 201-203
- [3] 黄映辉, 李冠宇. 物联网: 标志性特征与模型描述[J]. 计算机科学, 2011, 38(10A): 4-6, 17
- [4] 赵志军, 沈强, 唐晖, 等. 物联网架构和智能信息处理理论与关键技术[J]. 计算机科学, 2011, 38(8): 1-8
- [5] 童恩栋, 沈强, 雷军, 等. 物联网情景感知技术研究[J]. 计算机科学, 2011, 38(4): 9-14+20
- [6] 何明, 陈国华, 梁文辉, 等. 物联网环境下云数据存储安全及隐私保护策略研究[J]. 计算机科学, 2012, 39(5): 62-65, 90
- [7] 李玲娟. IoT的数据管理与智能处理[J]. 中兴通讯技术, 2011, 17(1): 38-41
- [8] 胡迎新, 马新娜, 郑丽娟. 物联网数据管理研究[J]. 物联网技术, 2014, 4: 79-82
- [9] 丁治明, 高需. 面向物联网海量传感器采样数据管理的数据库集群系统框架[J]. 计算机学报, 2012, 35(6): 1175-1191
- [10] IBM 联邦数据库技术[OL]. <http://www.ibm.com/developer-works/cn/data/library/techarticles/0203haas/0203haas.html>