



(12)发明专利申请

(10)申请公布号 CN 108108423 A

(43)申请公布日 2018.06.01

(21)申请号 201711352758.1

(22)申请日 2017.12.15

(71)申请人 吉旗(成都)科技有限公司

地址 610000 四川省成都市天府新区天府
大道南段846号

(72)发明人 吴少东 阎早华 张禄宇

(74)专利代理机构 北京权智天下知识产权代理
事务所(普通合伙) 11638

代理人 刘玉欣

(51)Int.Cl.

G06F 17/30(2006.01)

权利要求书1页 说明书3页

(54)发明名称

一种流式处理物联网数据的方法

(57)摘要

本发明公开了一种流式处理物联网数据的方法,包括原始数据收集,脏数据过滤,机构数据补全,省市区信息补全,经纬度运算,本发明在实际运用中,会实时接收数据,将查询数据时拼凑数据所占用时间,分布在每一条数据入库上,大大减少了查询耗时,提高了查询效率,减少了第三方服务的瞬时压力;本发明消除了调用方在使用数据时拼凑数据问题,此外,本发明通过流队列中数据之间的流转提高了各个部分之间的容错性。

1.一种流式处理物联网数据的方法,包括原始数据收集,脏数据过滤,机构数据补全,省市区信息补全,经纬度运算,其特征在于:包括以下步骤:

A、由汇通天下GSP上传原始报文,报文经解析之后,生成原始数据存储并发送到流队列中;

B、通过定义的任务获取流队列中的数据,并进行调用第三方服务,用以补全该条数据的机构信息,处理完毕后,流经下一个任务;

C、依次经过数据收集,脏数据处理,机构数据补全,省市区信息补全,经纬度运算等五个任务;

D、最后,数据将流入特定存储任务,存储于非关系型数据库以供查询。

2.根据权利要求1所述的一种流式处理物联网数据的方法,其特征在于:所述步骤C中脏数据处理方法包括以下步骤:

A、根据样例数据和样例数据的格式,定义数据格式,采集第一数据;

B、根据数据格式将采集的第一数据进行过滤和清洗,剔除脏数据,得到第二数据,并存储第二数据;

C、提取第二数据并按照业务逻辑进行数据处理;将第二数据进行数据转换;读取转换后的第二数据。

一种流式处理物联网数据的方法

技术领域

[0001] 本发明涉及数据处理技术领域,具体为一种流式处理物联网数据的方法。

背景技术

[0002] 物联网是通过各种信息传感设备,如传感器、射频识别(RFID)技术、全球定位系统、红外线感应器、激光扫描器、气体感应器等各种装置与技术,实时采集任何需要监控、连接、互动的物体或过程,采集其声、光、热、电、力学、化学、生物、位置等各种需要的信息,与互联网结合形成的一个巨大网络。其目的是实现物与物、物与人,所有的物品与网络的连接,方便识别、管理和控制。联网是新一代信息网络技术的高度集成和综合运用,是新一轮产业革命的重要方向和推动力量,对于培育新的经济增长点、推动产业结构转型升级、提升社会管理和公共服务的效率和水平具有重要意义。发展物联网必须遵循产业发展规律,正确处理好市场与政府、全局与局部、创新与合作、发展与安全的关系。要按照“需求牵引、重点跨越、支撑发展、引领未来”的原则,着力突破核心芯片、智能传感器等一批核心关键技术;着力在工业、农业、节能环保、商贸流通、能源交通、社会事业、城市管理、安全生产等领域,开展物联网应用示范和规模化应用;着力统筹推动物联网整个产业链协调发展,形成上下游联动、共同促进的良好格局;着力加强物联网安全保障技术、产品研发和法律法规制度建设,提升信息安全保障能力;着力建立健全多层次多类型的人才培养体系,加强物联网人才队伍建设。

[0003] 物联网在运行过程中产生较多的物联网数据,现有领域中均将数据存储在关系数据库中,在使用数据时由调用方去查询不同数据库补全信息,这就造成了调用方在使用数据时的复杂性。

发明内容

[0004] 本发明的目的在于提供一种流式处理物联网数据的方法,以解决上述背景技术中提出的问题。

[0005] 为实现上述目的,本发明提供如下技术方案:一种流式处理物联网数据的方法,包括原始数据收集,脏数据过滤,机构数据补全,省市区信息补全,经纬度运算,包括以下步骤:

[0006] A、由汇通天下GSP上传原始报文,报文经解析之后,生成原始数据存储并发送到流队列中;

[0007] B、通过定义的任务获取流队列中的数据,并进行调用第三方服务,用以补全该条数据的机构信息,处理完毕后,流经下一个任务;

[0008] C、依次经过数据收集,脏数据处理,机构数据补全,省市区信息补全,经纬度运算等五个任务;

[0009] D、最后,数据将流入特定存储任务,存储于非关系型数据库以供查询。

[0010] 优选的,所述步骤C中脏数据处理方法包括以下步骤:

- [0011] A、根据样例数据和样例数据的格式,定义数据格式,采集第一数据;
- [0012] B、根据数据格式将采集的第一数据进行过滤和清洗,剔除脏数据,得到第二数据,并存储第二数据;
- [0013] C、提取第二数据并按照业务逻辑进行数据处理;将第二数据进行数据转换;读取转换后的第二数据。
- [0014] 与现有技术相比,本发明的有益效果是:本发明公开的数据处理方法,在实际运用中,会实时接收数据,将查询数据时拼凑数据所占用时间,分布在每一条数据入库上,大大减少了查询耗时,提高了查询效率,减少了第三方服务的瞬时压力;本发明消除了调用方在使用数据时拼凑数据问题,此外,本发明通过流队列中数据之间的流转提高了各个部分之间的容错性;此外,本发明采用的脏数据处理方法能够对采集的数据进行过滤和清洗,剔除脏数据,提高了数据收集的有效性和数据处理的效率。

具体实施方式

[0015] 下面对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0016] 本发明提供如下技术方案:一种流式处理物联网数据的方法,包括原始数据收集,脏数据过滤,机构数据补全,省市区信息补全,经纬度运算,包括以下步骤:

[0017] A、由汇通天下GSP上传原始报文,报文经解析之后,生成原始数据存储并发送到流队列中;

[0018] B、通过定义的任务获取流队列中的数据,并进行调用第三方服务,用以补全该条数据的机构信息,处理完毕后,流经下一个任务;

[0019] C、依次经过数据收集,脏数据处理,机构数据补全,省市区信息补全,经纬度运算等五个任务;

[0020] D、最后,数据将流入特定存储任务,存储于非关系型数据库以供查询。

[0021] 本发明中,步骤C中脏数据处理方法包括以下步骤:

[0022] A、根据样例数据和样例数据的格式,定义数据格式,采集第一数据;

[0023] B、根据数据格式将采集的第一数据进行过滤和清洗,剔除脏数据,得到第二数据,并存储第二数据;

[0024] C、提取第二数据并按照业务逻辑进行数据处理;将第二数据进行数据转换;读取转换后的第二数据。

[0025] 本发明公开的数据处理方法,在实际运用中,会实时接收数据,将查询数据时拼凑数据所占用时间,分布在每一条数据入库上,大大减少了查询耗时,提高了查询效率,减少了第三方服务的瞬时压力;本发明消除了调用方在使用数据时拼凑数据问题,此外,本发明通过流队列中数据之间的流转提高了各个部分之间的容错性;此外,本发明采用的脏数据处理方法能够对采集的数据进行过滤和清洗,剔除脏数据,提高了数据收集的有效性和数据处理的效率。

[0026] 尽管已经示出和描述了本发明的实施例,对于本领域的普通技术人员而言,可以

理解在不脱离本发明的原理和精神的情况下可以对这些实施例进行多种变化、修改、替换和变型，本发明的范围由所附权利要求及其等同物限定。