



(12) 发明专利

(10) 授权公告号 CN 113360722 B

(45) 授权公告日 2022. 08. 09

(21) 申请号 202110711323.1

G06F 16/28 (2019.01)

(22) 申请日 2021.06.25

G06F 11/07 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113360722 A

(56) 对比文件

CN 111709527 A, 2020.09.25

CN 111737033 A, 2020.10.02

CN 112559237 A, 2021.03.26

CN 111858123 A, 2020.10.30

CN 112181960 A, 2021.01.05

CN 109905436 A, 2019.06.18

CN 112152830 A, 2020.12.29

CN 112416645 A, 2021.02.26

WO 2021104269 A1, 2021.06.03

US 2012005533 A1, 2012.01.05

刘瑞宏等. 基于知识图谱的智能故障诊断研究.《邮电设计技术》.2020,

(43) 申请公布日 2021.09.07

(73) 专利权人 杭州优云软件有限公司

地址 311305 浙江省杭州市临安区青山湖街道大园路958号科创大楼B座5楼505室

专利权人 北京广通优云科技股份有限公司

(72) 发明人 丁锐 王晓光 徐育毅

(74) 专利代理机构 杭州九洲专利事务所有限公司 33101

专利代理师 陈继亮

审查员 王芬

(51) Int. Cl.

G06F 16/901 (2019.01)

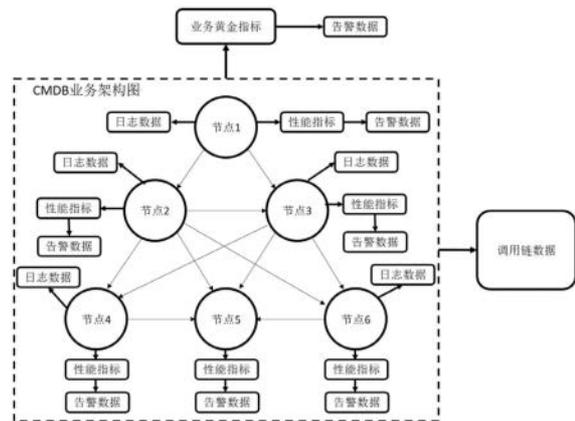
权利要求书2页 说明书9页 附图6页

(54) 发明名称

一种基于多维数据图谱的故障根因定位方法及系统

(57) 摘要

本发明公开了一种基于多维数据图谱的故障根因定位方法及系统,首先将多种维度数据进行综合分析后构建多维数据图谱,每种维度数据表示为数据图谱中的一个节点,多维数据图谱的构建方法中包含有关键性的索引映射关系节点,针对不同维度的数据建立不同的存储索引,构建成索引映射关系节点;然后基于该多维数据图谱,通过多种故障根因定位算法模式,确保在某一种或几种数据维度缺失时,能够正常进行故障的根因定位分析。本发明提出的方法和系统可以大幅缩短系统中断后的平均故障恢复时间,还可以减少对人工故障诊断和排查经验的依赖,还可以后续对接故障自动恢复功能,可以在故障发生时自动进行故障排查和故障恢复。



1. 一种基于多维数据图谱的故障根因定位方法,其特征在于:首先将多种维度数据进行综合分析后构建多维数据图谱,每种维度数据表示为数据图谱中的一个节点,多维数据图谱的构建方法中包含有关键性的索引映射关系节点,针对不同维度的数据建立不同的存储索引,并将索引信息通过映射关系连接起来,构建成索引映射关系节点,后续基于多维数据图谱的故障根因定位算法会首先使用该索引映射关系节点;然后基于该多维数据图谱,通过多种故障根因定位算法模式,确保在某一种或几种数据维度缺失时,能够正常进行故障的根因定位分析;

所述多维数据包括:某业务系统运行的业务黄金指标数据,支撑该业务系统的CMDB业务架构数据,该业务系统内部不同节点所依赖的多种调用链数据,反映该业务系统中节点运行时硬件或软件状态的性能指标数据,基于该性能指标或业务黄金指标异常时所产生的告警数据,记录该业务系统中各节点运行时中间信息的日志数据;

所述多维数据图谱的构建方法是:首先使用多种不同数据库联合存储然后再建立索引映射关系节点的构建方法,对于日志数据,放置在ElasticSearch中单独建立一个索引来存储,而性能指标数据在ElasticSearch中单独建立一个索引来存储,告警数据存储在多种关系型数据库中或存储在多种非关系型数据库中,然后根据所存储的数据库类型,数据库地址,端口号建立每一个节点的数据映射关系,构建出索引映射关系节点,后续能够基于这些索引映射关系节点来快速检索和分析多维数据图谱;

所述的故障根因定位算法模式为:以业务某种指标为故障根因定位的入口,首先使用动态基线算法诊断某种指标,判断是否发生异常,如果某种指标异常,则判定发生了故障,在故障发生时,首先对调用链数据使用调用链异常检测算法,定位出是具体哪个节点发生了故障;定位到某一节点发生故障,则聚焦到该节点上,获取该节点的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常;然后获取该节点的日志数据,使用日志异常检测算法来判断是否有日志发生异常,具体哪些日志异常;最后将汇总后的异常节点,异常指标,异常日志数据,异常指标对应的告警数据,异常日志对应的告警数据都保存到数据库中,前端页面获取这部分数据后展示;

在业务某种指标缺失的情况下,使用多维数据图谱进行故障根因定位,首先在业务某种指标缺失时,由于无法确定何时会发生故障,故而采用定时循环遍历调用链数据,来确定发生故障的具体时间和故障发生的节点位置;遍历时定位到某一节点发生故障,则聚焦到该节点上,获取该节点的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,然后获取该节点的日志数据,使用日志异常检测算法来判断是否有日志发生异常,具体哪些日志异常,最后将汇总后的异常节点,异常指标,异常日志数据,异常指标对应的告警数据,异常日志对应的告警数据都保存到数据库,起到对故障进行根因定位的效果。

2. 根据权利要求1所述的基于多维数据图谱的故障根因定位方法,其特征在于:综合分析方法包括:对多维度数据进行实时采集并接入方法,包括从Kafka数据源接入实时数据,以Restful API的方式采集到的实时数据,对数据源轮询后得到的多种实时数据;对多维度数据进行相关性分析,相关性分析方法包括基于物理拓扑结构的相关性分析方法,基于数值波动相似性的相关性分析方法,基于业务逻辑关联的相关性分析方法;还包括对多维数据进行自动规范化分析方法,对某些有数据缺失的,采集数据不规范的,数据格式不统一的多维数据进行规范化的,统一的数据规范化分析。

3. 根据权利要求1所述的基于多维数据图谱的故障根因定位方法,其特征在于:所述多维数据图谱的构建方法采用基于关系型数据库建立索引映射关系节点的构建方法,将多维度数据分别存储在不同数据表中,或存储到同一个数据表中,但以特殊字段来区分;然后针对不同表的地址和表名称,结合数据之间的对应关系,建立索引映射关系节点。

4. 根据权利要求1所述的基于多维数据图谱的故障根因定位方法,其特征在于:所述的故障根因定位算法模式为:以业务某种指标作为故障根因定位的入口,首先使用动态基线算法判断某种指标是否发生异常,如果出现异常,则判定发生故障,然后,对该业务系统中所有的节点进行循环遍历,使用指标异常检测算法依次诊断每一个节点中的所有指标数据,对每个节点中的指标数据异常情况得到概率值,做平均处理后,取概率最大的节点作为异常节点,并将这些指标的异常概率从大到小排序,获取前N个指标作为根因指标,同时对该异常节点的日志数据进行分析,判断是否发生异常,发生异常时一起将异常日志信息作为根因保存到数据库中。

5. 根据权利要求1所述的基于多维数据图谱的故障根因定位方法,其特征在于:在业务某种指标和日志数据同时缺失的情况下,使用多维数据图谱进行故障根因定位,首先在业务某种指标缺失时,由于无法确定何时会发生故障,故而采用定时循环遍历调用链数据,来确定发生故障的具体时间和故障发生的节点位置;遍历时定位到某一节点发生故障,则聚焦到该节点上,获取该节点的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,最后将汇总后的异常节点,异常指标,异常指标对应的告警数据,都保存到数据库,起到对故障进行根因定位的效果。

6. 一种采用如权利要求1-5任意一项基于多维数据图谱的故障根因定位方法的系统,其特征在于:该系统包括有:多维数据图谱模块,根因定位算法组模块,算法模式调度器模块,数据缓存模块,界面展示模块,多种模块有机结合后确保系统稳定运行,并及时有效的在某业务系统发生故障后进行故障的根因定位。

一种基于多维数据图谱的故障根因定位方法及系统

技术领域

[0001] 本发明涉及IT运维系统中的故障分析技术领域,主要是一种基于多维数据图谱的故障根因定位方法及系统。

背景技术

[0002] 随着数据中心基础设施的快速扩张,IT运维所管理的主机规模不断扩大,IT系统架构越来越复杂,IT运维响应时效要求越来越高,云时代下IT运维工作面临着“系统冗余告警多,故障根因定位难”,“故障发生时,人工难以进行快速排查,或人工排查故障的准确性难以保证”等诸多问题和痛点。随着主机规模不断扩大,云时代架构复杂度不断提升,IT运维中发生故障的根本原因也不断的多样化,复杂化,比如运维中经常会存在一些由于版本变更、业务变化、代码逻辑或网络波动等带来的各种故障,尤其是系统较为庞大、复杂时,进行故障的根因分析就显得尤为困难。一旦某业务系统出现问题,将给企业带来重大的业务影响,造成巨大的业务损失。

[0003] IT运维中故障的根因分析就是当某业务系统发生故障时,从各种不同的数据维度进行汇聚分析,比如所涉及业务系统包含的反映系统软硬件运行状态的多种性能指标数据,记录业务系统运行时的各种中间信息的日志数据,反映系统中不同网元之间相互调用关系的调用链数据,反映某业务系统是否发生异常的多种业务黄金指标等数据,由于引起故障的原因复杂,涉及的系统组件众多,包含的数据多样,每种维度数据的量非常大等原因,导致人工排查业务系统故障非常困难,人工定位故障的根因分析耗时长,严重依赖人工运维经验等。

[0004] 故而,为了能够自动化,智能化的对IT系统中的故障进行自动诊断和根因定位,目前有一些专利试图引入人工智能算法来进行故障的根因定位,比如专利CN202011353855提出了一种应用故障根因定位方法及系统,其主要是通过目标应用对应的各个节点的指标数据组,应用预设的基线拟合模型,评分准则确定根因定位的评分,根据该评分大小来进一步确定所述目标应用是否为故障根因节点。但该专利仅仅使用的是监控指标数据,配合监控报警数据,使用的是预设的动态基线拟合模型,这种算法仅仅依靠一种或两种维度的数据,仅仅使用动态基线拟合来进行根因定位,由于数据维度太少故而忽略了不同节点之间的交互关系,难以全面的,准确的进行故障根因定位,特别是在节点调用关系复杂,节点数据众多的情况下难以得到理想的根因定位结果。

[0005] 针对以上问题,专利CN202011554734公开了一种基于知识图谱的云原生系统故障分析方法,该方法首先获取云原生系统中的原始数据,并基于该原始数据构建知识图谱,得到图数据,然后使用异常检测模型对该图数据进行异常讲的,得到异常节点,然后计算异常节点和异常节点对应的副本节点的相似度,基于该相似度进行故障根因定位。但这种方法仅仅使用了是文本型数据来构建知识图谱,比如对告警数据,日志数据进行关系抽取和实体识别,然后构建知识图谱,没有用到云原生系统中非常重要的反映节点之间调用关系的调用链数据,反映节点硬件或软件运行状态的性能指标数据,故而该方法所使用的数据维

度太少,难以全面地在故障发生时进行全局维度地故障根因分析。

发明内容

[0006] 针对目前IT运维系统中故障根因定位所用数据维度太少,难以全面进行根因分析,所用AI算法单一导致故障根因定位准确性低等缺点,本发明公开一种基于多维数据图谱的故障根因定位方法及系统。

[0007] 本发明的目的是通过如下技术方案来完成的。一种基于多维数据图谱的故障根因定位方法,首先将多种维度数据进行综合分析后构建多维数据图谱,每种维度数据表示为数据图谱中的一个节点,多维数据图谱的构建方法中包含有关键性的索引映射关系节点,针对不同维度的数据建立不同的存储索引,并将索引信息通过映射关系连接起来,构建成索引映射关系节点,后续基于多维数据图谱的故障根因定位算法会首先使用该索引映射关系节点;然后基于该多维数据图谱,通过多种故障根因定位算法模式,确保在某一种或几种数据维度缺失时,能够正常进行故障的根因定位分析。

[0008] 所述多维数据包括但不限于:某业务系统运行的业务黄金指标数据,支撑该业务系统的CMDB业务架构数据,该业务系统内部不同节点所依赖的多种调用链数据,反映该业务系统中节点运行时硬件或软件状态的性能指标数据,基于该性能指标或业务黄金指标异常时所产生的告警数据,记录该业务系统中各节点运行时中间信息的日志数据。

[0009] 所述的综合分析方法包括但不限于:对多维度数据进行实时采集并接入方法,包括但不限于从Kafka数据源接入实时数据,以Restful API的方式采集到的实时数据,对数据源轮询后得到的多种实时数据;对多维度数据进行相关性分析,该相关性分析方法包括但不限于基于物理拓扑结构的相关性分析方法,基于数值波动相似性的相关性分析方法,基于业务逻辑关联的相关性分析方法;还包括对多维数据进行自动规范化分析方法,对某些有数据缺失的,采集数据不规范的,数据格式不统一的多维数据进行规范化的,统一的数据规范化分析。

[0010] 所述多维数据图谱的构建方法采用首先基于关系型数据库再建立索引映射关系节点的构建方法,将多维度数据分别存储在不同数据表中,或存储到同一个数据表中,但以特殊字段来区分;然后针对不同表的地址和表名称,结合数据之间的对应关系,建立索引映射关系节点。

[0011] 所述多维数据图谱的构建方法是:首先使用多种不同数据库联合存储然后再建立索引映射关系节点的构建方法,对于日志数据,放置在ElasticSearch中单独建立一个索引来存储,而性能指标数据在ElasticSearch (ES) 中单独建立一个索引来存储,告警数据存储多种关系型数据库中或存储在多种非关系型数据库中,然后根据所存储的数据库类型,数据库地址,端口号建立每一个节点的数据映射关系,构建出索引映射关系节点,后续能够基于这些索引映射关系接待了来快速检索和分析多维数据图谱。

[0012] 所述的故障根因定位算法模式为:以业务某种指标为故障根因定位的入口,首先使用动态基线算法诊断某种指标,判断是否发生异常,如果某种指标异常,则判定发生了故障,在故障发生时,首先对调用链数据使用调用链异常检测算法,定位出是具体哪个节点发生了故障,比如定位到节点3发生故障,则更进一步的,聚焦到节点3上,获取节点3的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,然后获取节点3的日志数

据,使用日志异常检测算法来判断是否有日志发生异常,具体哪些日志异常,最后将汇总后的异常节点,异常指标,异常日志数据,异常指标对应的告警数据,异常日志对应的告警数据都保存到数据库中,前端页面获取这部分数据后展示。

[0013] 所述的故障根因定位算法模式为:以业务某种指标作为故障根因定位的入口,首先使用动态基线算法判断某种指标是否发生异常,如果出现异常,则判定发生故障,然后,对该业务系统中所有的节点进行循环遍历,使用指标异常检测算法依次诊断每一个节点中的所有指标数据,对每个节点中的指标数据异常情况得到概率值,做平均处理后,取概率最大的节点作为异常节点,并将这些指标的异常概率从大到小排序,获取前N个指标作为根因指标,同时对该异常节点的日志数据进行分析,判断是否发生异常,发生异常时一起将异常日志信息作为根因保存到数据库中。

[0014] 所述的故障根因定位算法模式为:在业务某种指标缺失的情况下,仍然使用本发明公开的多维数据图谱进行故障根因定位,首先在业务某种指标缺失时,由于无法确定何时会发生故障,故而采用定时循环遍历调用链数据,来确定发生故障的具体时间和故障发生的节点位置,比如遍历时定位到节点3发生故障,则更进一步的,聚焦到节点3上,获取节点3的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,然后获取节点3的日志数据,使用日志异常检测算法来判断是否有日志发生异常,具体哪些日志异常,最后将汇总后的异常节点,异常指标,异常日志数据,异常指标对应的告警数据,异常日志对应的告警数据都保存到数据库,起到对故障进行根因定位的效果。

[0015] 所述的故障根因定位算法模式为:在业务某种指标和日志数据同时缺失的情况下,使用本发明公开的多维数据图谱进行故障根因定位,首先在业务某种指标缺失时,由于无法确定何时会发生故障,故而采用定时循环遍历调用链数据,来确定发生故障的具体时间和故障发生的节点位置,比如遍历时定位到节点3发生故障,则更进一步的,聚焦到节点3上,获取节点3的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,最后将汇总后的异常节点,异常指标,异常指标对应的告警数据,都保存到数据库,起到对故障进行根因定位的效果。

[0016] 本发明同时公开了一种基于多维数据图谱的故障根因定位系统,该系统包括有:多维数据图谱模块,根因定位算法组模块,算法模式调度器模块,数据缓存模块,界面展示模块,多种模块有机结合后可以确保系统稳定运行,并及时有效的在某业务系统发生故障后进行故障的根因定位。

[0017] 本发明的有益效果为:

[0018] 针对IT系统中故障排查困难,故障根因难以定位,严重依赖运维人员的人工经验等问题,本发明创新型的提出了一种基于多维数据图谱的,针对IT运维领域的故障根因定位算法,并基于该算法构建了一种故障根因定位系统。基于多维数据图谱的故障根因定位算法首先会收集历史发生的针对某种故障的多种维度的数据,包括但不限于:某业务系统的业务黄金指标,性能指标,调用链数据,日志数据,告警数据,CMDB数据;并基于这些数据建立多维数据图谱,梳理出不同数据之间的依赖关系。

[0019] 基于该多维数据图谱,本发明创建了一种智能定位方法和系统,可以对IT运维系统中的业务系统进行诊断分析,本发明提出的故障根因定位算法可以在无需人为干预的情况下,智能化的,自动的进行故障诊断,更进一步的,在故障发生时,可以用AI算法实现故障

根因的自动定位,并给出定位后的故障根因结果。本发明提出的算法和系统可以大幅缩短系统中断后的平均故障恢复时间(MTTR),还可以减少对人工故障诊断和排查经验的依赖,本算法和系统还可以后续对接故障自动恢复功能,可以在故障发生时自动进行故障排查和故障恢复。

附图说明

- [0020] 图1为本发明的一种示例性的多维数据图谱样式;
- [0021] 图2为本发明的一种示例性的对某节点建立的索引映射关系节点部分视图;
- [0022] 图3为示例性的故障根因定位算法模式一的实施方案流程图。
- [0023] 图4为示例性的故障根因定位算法模式二的实施方案流程图。
- [0024] 图5为示例性的故障根因定位算法模式三的实施方案流程图。
- [0025] 图6为示例性的故障根因定位算法模式四的实施方案流程图。
- [0026] 图7为一种示例性的基于多维数据图谱的故障根因定位系统。

具体实施方式

[0027] 下面将结合附图对本发明做详细的介绍:

[0028] 本发明公开了一种基于多维数据图谱的故障根因定位方法及系统,首先将多种维度数据进行综合分析后构建多维数据图谱,每种维度数据表示为数据图谱中的一个节点,然后基于该多维数据图谱,公开了多种定位算法模式,确保在某一种或几种数据维度缺失时,可以正常进行故障的根因定位分析。基于所述多维数据图谱和所述多种定位算法模式,本发明还公开了一种故障根因定位系统,该系统包括有:多维数据图谱模块,根因定位算法组模块,算法模式调度器模块,数据缓存模块,界面展示模块等,多种模块有机结合后可以确保系统稳定运行,并及时有效的在某业务系统发生故障后进行故障的根因定位。

[0029] 本发明首先公开一种与某业务系统相关联的多维数据进行综合分析并构建多维数据图谱的方法,此处所指的综合分析方法包括但不限于:对多维度数据进行实时采集并接入方法,包括但不限于从Kafka数据源接入实时数据,以Restful API的方式采集到的实时数据,对数据源轮询后得到的多种实时数据等;对多维度数据进行相关性分析,该相关性分析方法包括但不限于基于物理拓扑结构的相关性分析方法,基于数值波动相似性的相关性分析方法,基于业务逻辑关联的相关性分析方法等;还包括对多维数据进行自动规范化分析方法,对某些有数据缺失的,采集数据不规范的,数据格式不统一的多维数据进行规范化的,统一的数据规范化分析等。

[0030] 本发明所述业务系统并不仅限于某种具体的应用系统,可以指代例如银行业务中的第三方快捷支付系统,手机银行支付系统等,但也可以指代其他多种基于云原生应用系统。

[0031] 本发明所述多维数据包括但不限于:某业务系统运行的业务黄金指标数据,支撑该业务系统的CMDB业务架构数据,该业务系统内部不同节点所依赖的多种调用链数据,反映该业务系统中节点运行时硬件或软件状态的性能指标数据,基于该性能指标或业务黄金指标异常时所产生的告警数据,记录该业务系统中各节点运行时中间信息的日志数据。

[0032] 本发明所述构建的多维数据图谱的一种组织形式为如下图所示,这仅仅是示例性

的说明,并不是限制性的认为本发明所述多维数据图谱就是图1所示样例。在该多维数据图谱中,仅仅示例性的标识出某业务系统具有6个节点,但节点数和节点之间的连接方式,连接方向等可以多样化。CMDDB业务架构图中的节点可以和性能指标数据,日志数据直接关联起来,而告警数据可以和性能指标数据,或者业务黄金指标数据关联起来,更具体的,在某个性能指标出现异常时,比如CPU的使用率指标超过固定阈值85时,会产生一条告警数据,进而告警数据可以与指标数据关联起来。更进一步的,告警数据还可以与日志数据关联起来,即当对日志数据进行异常检测并检测出日志出现异常时,就会产生一条日志异常相关的告警数据。

[0033] 本发明所述多维数据图谱的构建方法中包含有关键性的索引映射关系节点,其主要作用是,针对不同维度的数据建立不同的存储索引,并将索引信息通过映射关系连接起来,构建成索引映射关系节点,后续基于多维数据图谱的故障根因定位算法会首先使用该索引映射关系节点。

[0034] 作为示例性的,多维数据图谱的构建方法可以采用首先基于关系型数据库再建立索引映射关系节点的构建方法,比如在MySQL,Oracle等数据库中,可以将多维度数据分别存储在不同数据表中,也可以存储到同一个数据表中,但以特殊字段来区分。比如节点2的日志数据单独存放在一张表(eg:node2_weblogic_log_20200103)中,性能指标数据存放在另外一张表(eg:node2_metric_20200103)中,告警数据存放在另外一张表(eg:node2_alert_20200103)中,然后针对不同表的地址和表名称,结合数据之间的对应关系,建立索引映射关系节点,比如此处所建立的索引映射关系节点一部分内容可以表示为如下图所示。

[0035] 作为另一种示例性的,本发明所述多维数据图谱的构建方法还可以是:首先使用多种不同数据库联合存储然后再建立索引映射关系节点的构建方法,比如对于日志数据,通常数据量非常大,可以放置在ElasticSearch (ES) 中单独建立一个索引来存储,而性能指标数据也可以在ElasticSearch (ES) 中单独建立一个索引来存储,告警数据可以存储在多种关系型数据库中,比如MySQL,Oracle等数据库中,也可以存储在多种非关系型数据库中,比如MongoDB, NoSQL数据库中,然后根据所存储的数据库类型,数据库地址,端口号等建立每一个节点的数据映射关系,构建出索引映射关系节点,后续可以基于这些索引映射关系接待了来快速检索和分析多维数据图谱。

[0036] 基于所述多维数据图谱,本发明公开了多种故障根因定位算法模式,以应对多种不同应用场景下的业务需求,每种故障根因定位算法均可以单独用于故障诊断,故障根因分析,故障根因结果展示环节。

[0037] 更具体地,作为示例性的,故障根因定位算法模式一的实施方案为:以某业务系统为例,以业务黄金指标为故障根因定位的入口,首先使用动态基线算法诊断黄金指标,判断是否发生异常,如果黄金指标异常,则判定发生了故障,在故障发生时,首先对调用链数据使用调用链异常检测算法,定位出是具体哪个节点发生了故障,比如定位到节点3发生故障,则更进一步的,聚焦到节点3上,获取节点3的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,然后获取节点3的日志数据,使用日志异常检测算法来判断是否有日志发生异常,具体哪些日志异常,最后将汇总后的异常节点,异常指标,异常日志数据,异常指标对应的告警数据,异常日志对应的告警数据都保存到数据库中,前端页面获取

这部分数据后展示。此处所述故障根因定位算法模式一的流程图如图3所示。

[0038] 作为示例性的,故障根因定位算法模式二的实施方案为:以某业务系统为例,以业务黄金指标作为故障根因定位的入口,首先使用动态基线算法判断黄金指标是否发生异常,如果出现异常,则判定发生故障,然后,对该业务系统中所有的节点进行循环遍历,使用指标异常检测算法依次诊断每一个节点中的所有指标数据,对每个节点中的指标数据异常情况得到概率值,做平均处理后,取概率最大的节点作为异常节点,并将这些指标的异常概率从大到小排序,获取前N个(可以人为设置N,比如N=5)指标作为根因指标,同时对该异常节点的日志数据进行分析,判断是否发生异常,发生异常时一起将异常日志信息作为根因保存到数据库中。此处所述故障根因定位算法模式二的流程图如图4所示。

[0039] 需要注意的是,故障根因定位算法模式二是针对某些业务系统中没有调用链数据的情况下进行的根因定位流程,为了增加本发明的普遍适用性,本发明考虑到在某些应用系统中,无法提供关于该系统的调用链数据,故而在这种情况下,可以采用类似于本发明公开的故障根因定位算法模式二的方式,使用本发明公开的多维数据图谱,仍然能够进行某业务系统的故障根因定位方法及其系统。

[0040] 作为示例性的,故障根因定位算法模式三的实施方案为:以某业务系统为例,在业务黄金指标缺失的情况下,仍然可以使用本发明公开的多维数据图谱进行故障根因定位,作为示例性说明的,首先在业务黄金指标缺失时,由于无法确定何时会发生故障,故而本方案采用定时循环遍历调用链数据,来确定发生故障的具体时间和故障发生的节点位置,比如遍历时定位到节点3发生故障,则更进一步的,聚焦到节点3上,获取节点3的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,然后获取节点3的日志数据,使用日志异常检测算法来判断是否有日志发生异常,具体哪些日志异常,最后将汇总后的异常节点,异常指标,异常日志数据,异常指标对应的告警数据,异常日志对应的告警数据都保存到数据库,起到对故障进行根因定位的效果。此处所述故障根因定位算法模式三的流程如图5所示。

[0041] 需要注意的是,故障根因定位算法模式三是在多维数据图谱中缺失业务黄金指标数据的情况下采用的一种算法流程,但在多维数据图谱中存在业务黄金指标数据时,仍然可以人为弃用该部分数据,起到选用部分多维数据图谱的组件进行故障根因定位的效果。

[0042] 更需要注意的是,基于本发明所公开的多维数据图谱的部分维度数据,虽然可以通过类似故障根因定位算法模式三来完成故障的根因定位流程,但缺少业务黄金指标数据维度时,采用定时循环遍历调用链数据的方式会导致系统计算量上升,导致该算法模式需要消耗更多的计算资源,对系统性能要求更高,而且同时得到的故障根因定位算法在时效性上有所降低。

[0043] 作为示例性的,故障根因定位算法模式四的实施方案为:以某业务系统为例,在业务黄金指标和日志数据同时缺失的情况下,仍然可以使用本发明公开的多维数据图谱进行故障根因定位,作为一种示例性说明,首先在业务黄金指标缺失时,由于无法确定何时会发生故障,故而本方案采用定时循环遍历调用链数据,来确定发生故障的具体时间和故障发生的节点位置,比如遍历时定位到节点3发生故障,则更进一步的,聚焦到节点3上,获取节点3的所有性能指标数据,使用指标异常检测算法判断哪些指标发生了异常,最后将汇总后的异常节点,异常指标,异常指标对应的告警数据,都保存到数据库,起到对故障进行根因

定位的效果。此处所述故障根因定位算法模式四的流程图如图6所示。

[0044] 需要注意的是,基于本发明所公开的多维数据图谱的部分维度数据,虽然可以通过类似故障根因定位算法模式四来完成故障的根因定位流程,但缺乏围绕故障的业务黄金指标和日志数据,会使得该算法模式需要消耗更多的计算资源,对系统性能要求较高,同时缺乏日志维度数据,难以准确的定位到故障根因,故而这种算法模式得到的故障根因结果在时效性上有所降低,同时在准确性上也会下降。

[0045] 本发明所述故障根因定位算法模式中所提及的针对业务黄金指标的动态基线算法包括但不限于:对历史业务黄金指标进行统计学分析,包括同比量计算,环比量计算,时间窗口中指标波动率,指标最大值,最小值,中值,平均值等计算,通过多种特征和机器学习算法出来的指标的上下波动范围;对历史数据进行自回归线性拟合计算得到的上下波动范围;使用诸多机器学习算法(比如支持向量机SVM,lightGBM,XGboost算法等)对历史的某时间窗口内的指标数据进行预测分析,得到的未来预测值,并在预测值获取一定范围(例如:20%)所得到的上下波动范围;使用神经网络,自编码器算法,或者时序型循环神经网络所计算出来的指标在某时间段内的正常波动范围;使用人工经验,对单条或多条业务黄金指标进行分析后得到的不同时间段内的正常波动范围等。

[0046] 本发明所述故障根因定位算法模式中所提及的针对调用链数据的调用链异常检测算法包括但不限于:使用神经网络,自编码器算法,时序型循环神经网络算法所计算出来的调用链的正常范围,并判断实际调用链数据是否超出正常范围的异常检测算法;使用调用链中不同跨度间的耗时占比,调用链耗时统计分析,并基于统计进行的3-sigma算法,箱线图算法进行的异常检测;使用不同时间窗口内调用链个数的统计值,并基于统计值所进行的类似于孤立森林算法,类似于局部异常因子算法(LoF)的异常检测算法等。

[0047] 本发明所述故障根因定位算法模式中所提及的针对性能指标数据的指标异常检测算法包括但不限于:对性能指标数据进行切片处理,使用切片后的数据进行波动范围建模,确定正常范围的上下基准,将真实指标值与上下基准进行比对来判断指标是否异常的算法;对性能指标进行滑动窗口处理,获取一定窗口内部的指标数据,并对该窗口中数据进行统计值分析,比如最大值,最小值,中值,平均值,方差,同比值,环比值等,并基于这些特征建立机器学习模型,比如基于支持向量机SVM,基于随机森林RF,基于lightGBM,基于神经网络深度学习模型等多种机器学习算法所得到的异常检测结果;对一定时间段内的指标数据进行时序性建模,使用类似于循环神经网络,自编码器算法等对未来指标值进行预测,并判断真实值与预测值之间的差异水平,以此判断指标是否异常等。

[0048] 本发明所述故障根因定位算法模式中所提及的针对日志数据的日志异常检测算法包括但不限于:对实时产生的流式日志数据进行固定窗口取值处理,并对窗口内的日志数据进行分词处理,去除停用词,并统计关键字的出现频次,基于关键字和频次特征来判断日志是否发生异常的日志异常检测算法;对实时产生的日志数据进行滑动窗口处理,并对窗口内的日志数据进行日志模板分析,对日志模板进行编号并与历史日志数据中的日志模板进行匹配,判断新增日志,低频次日志,以及含有特殊关键字的日志等,在这些日志发生时判定为异常现象的异常检测算法;对实时产生的日志数据首先进行统一的树形结构算法分析,抽取出日志固有的文本模板,并对文本模板中的关键词进行统计分析,关键数值进行抽取建模,在关键数值过大或过小时判定日志异常的检测算法等。

[0049] 基于所述多维数据图谱和所述多种故障定位算法模式,可以建立多种故障根因定位系统,本发明还公开了一种示例性的故障根因定位系统,该系统包括有:多维数据图谱模块,根因定位算法组模块,算法模式调度器模块,数据缓存模块,界面展示模块等,多种模块有机结合后可以确保系统稳定运行,并及时有效的在某业务系统发生故障后进行故障的根因定位。本发明公开的一种示例性的基于多维数据图谱的故障根因定位系统如下图7所示。

[0050] 本发明所述根因定位系统中所包含的数据缓存模块的主要功能是:在真实的业务系统环境中,某业务系统的多维度数据都是实时地快速地产生,并快速的传递到后端进行消费的,由于故障根因定位算法需要一定时间,故而为了避免数据丢失,需要在此处设置数据缓存模块用于接收和暂时保存实时产生的多维度数据,此处的暂时保存并不做时间上的限制,可以是几秒,几分钟,几个小时,几天,甚至几个月的时间,数据的保存方式可以是暂时存储到主机的内存中,也可以是存储到其他主机中的专用数据库中,也可以是暂时存储在类似于Kafka等中间件中,还可以是存储在类似于RabbitMQ等消息队列中。保存的多维度数据可以按照一定的先后顺序来组织起来,避免数据错乱,也可以按照不同磁盘分区,不同数据库表名,不同文件系统索引名来区分。

[0051] 本发明所述根因定位系统中所包含的多维数据图谱模块的主要功能是:定时获取上游数据缓存模块中的多维度数据,并做初步数据处理,此处所述初步数据处理方式包括有:对连续型实时数据进行切片处理,对流式数据进行滑动窗口处理,对切片后或窗口中的数据进行基础统计分析,对日志,告警等文本类数据进行结构化处理,对非格式化的数据进行统一的格式化处理等。此处的多维数据图谱模块在此基础上还会将这些多维数据构建成多维数据图谱,并为每种维度的数据创建存储索引,针对不同的索引创建一一对应的关系映射,并将这种映射关系保存到专用的关系节点上,此处的存储索引会为后续的故障根因定位算法提供数据支撑。

[0052] 本发明所述根因定位系统中所包含的根因定位算法组模块的主要功能是:汇聚多种智能算法,比如包含针对业务黄金指标的动态基线算法,包含针对调用链数据的调用链异常检测算法,包含针对性能指标数据的指标异常检测算法,包含针对日志数据的日志异常检测算法等,所述根因定位算法组模块并不是简单的算法代码的堆叠,而是预设有多种相关调用逻辑,分别表示出在何种情况下运用何种检测算法,同时,所述根因定位算法组模块还包含有多种算法的参数配置信息,对于每种算法,都提供默认的算法参数配置,但需要注意的是,这些参数配置可以通过界面展示模块来调整,进而会影响算法的实现速率和达到精度效果。

[0053] 本发明所述根因定位系统中所包含的算法模式调度器模块的主要功能是:首先通过所述多维数据图谱模块中索引映射关系节点来判断当前系统所具备的数据维度是否全面,当前业务系统共接入了哪些维度的数据,数据是否符合一定规范,是否具有结构化存储方式等,所述算法模式调度器模块然后根据所接入的数据维度,对不同的维度数据选择根因定位算法组中的不同检测算法,起到根据不同来源数据调度不同检测算法模式的作用,作为示例性的,所述算法模式调度器模块在多维数据图谱中的数据维度比较全面是,会调度根因定位算法组模块中故障根因定位算法模式一相匹配的算法流程,另一种示例性的,针对多维数据图谱中缺失业务黄金指标数据时,会调度根因定位算法组模块中故障根因定位算法模式三相匹配的算法流程。

[0054] 本发明所述根因定位系统中所包含的界面展示模块的主要功能是：将原始的未经处理的多维度数据展示到界面中，方便运维人员查看原生数据变化趋势，以及在根因定位算法准确性难以达到要求时，方便运维人员根据人工经验进行专家会诊和人工排查。同时，将本发明所述根因定位方法的最终结果展示到界面上，方便运维人员查看定位结果，根据结果进行故障排查，故障修复等操作。另一方面，为了方便运维人员对算法进行调试，本界面展示模块还可以提供交互界面，方便运维人员针对根因定位算法组模块中输入不同算法的参数配置，并将该输入的参数配置保存到后端配置文件中，用于下次算法提升的参数选择。

[0055] 可以理解的是，对本领域技术人员来说，对本发明的技术方案及发明构思加以等同替换或改变都应属于本发明所附的权利要求的保护范围。

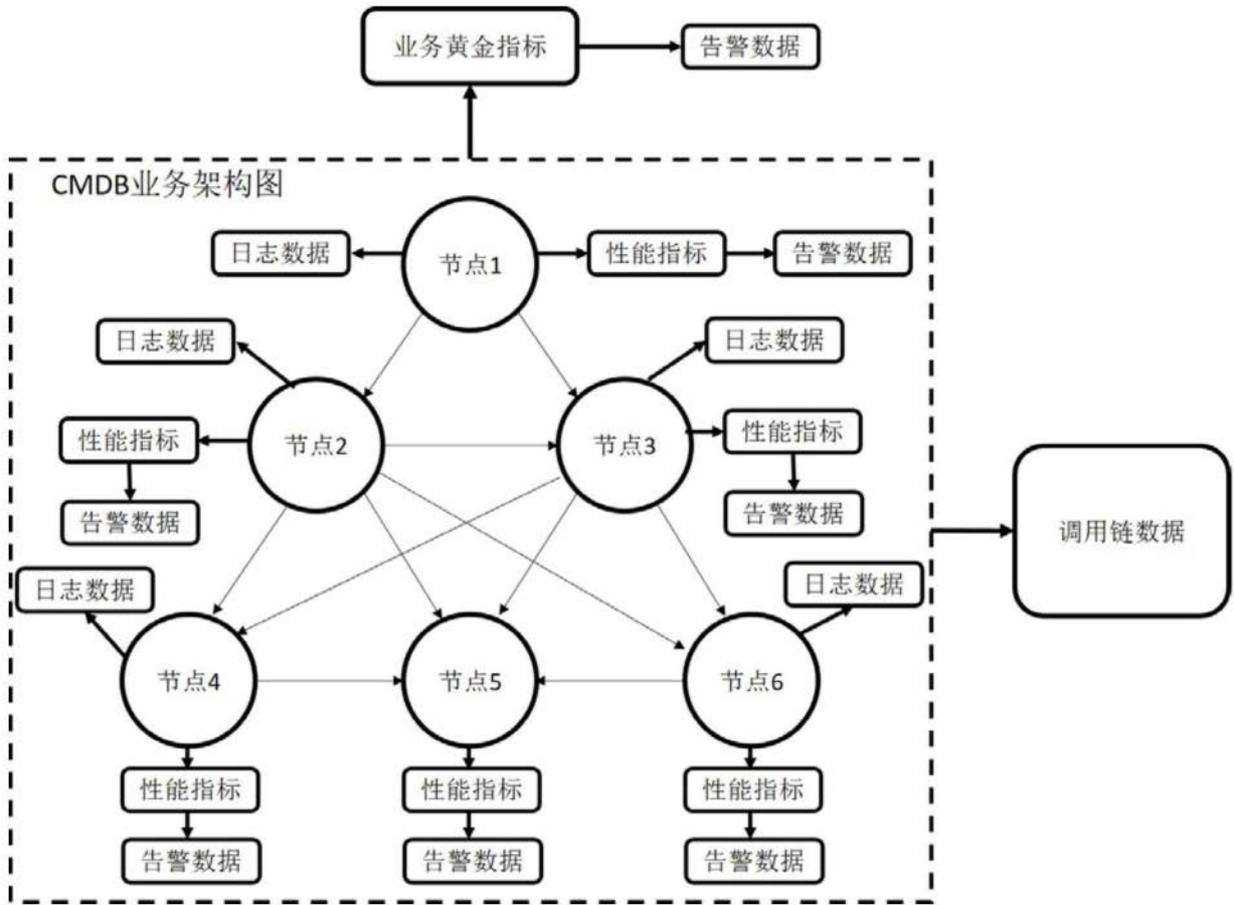


图1

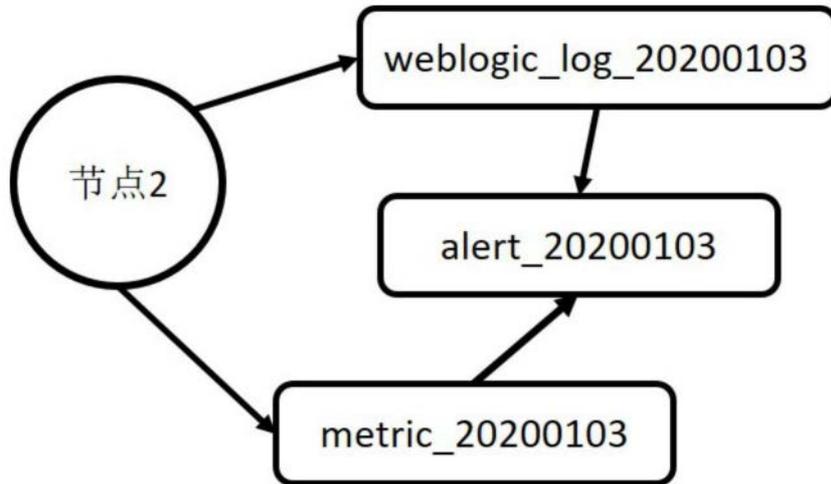


图2

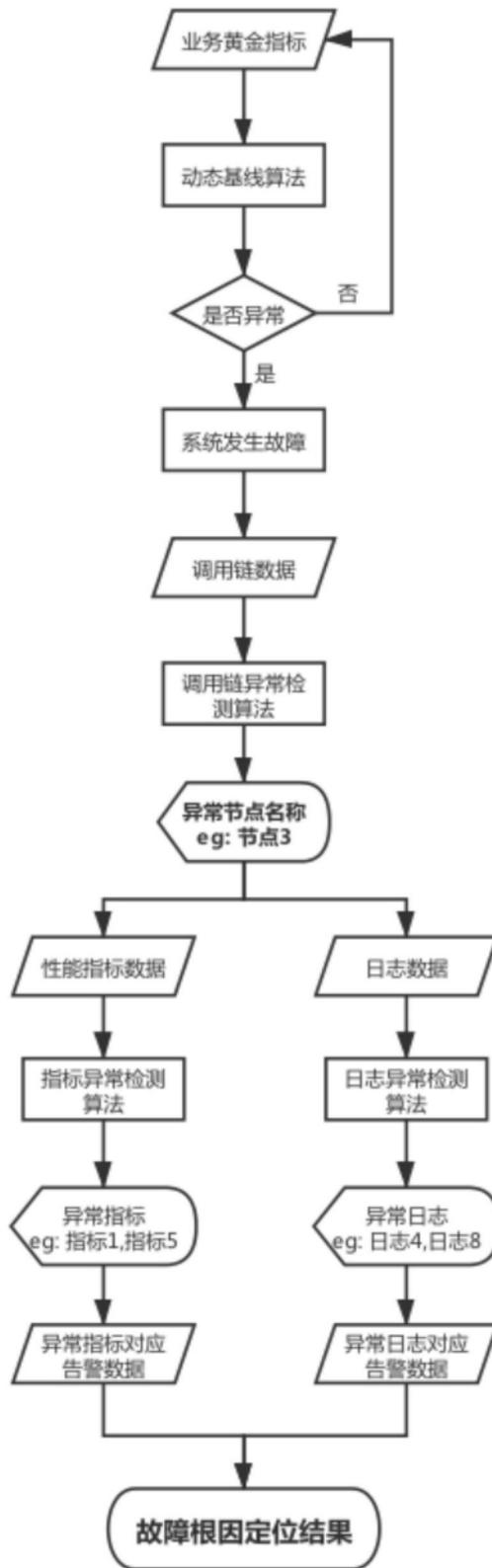


图3

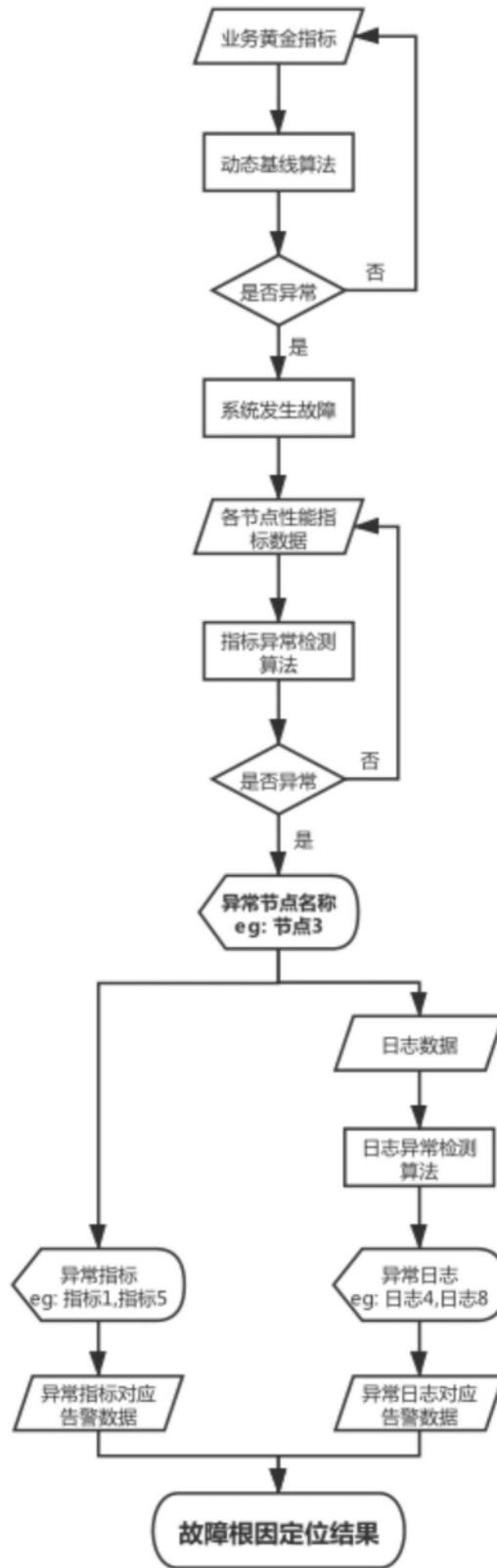


图4

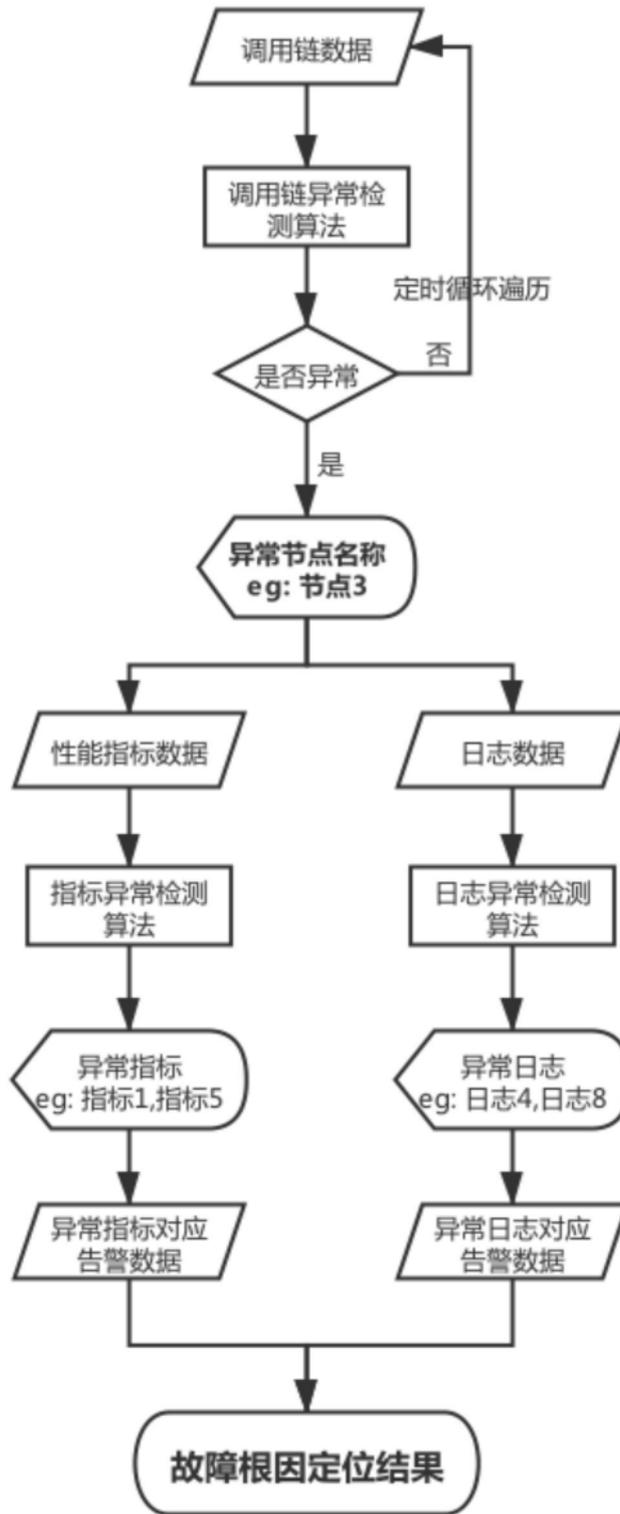


图5

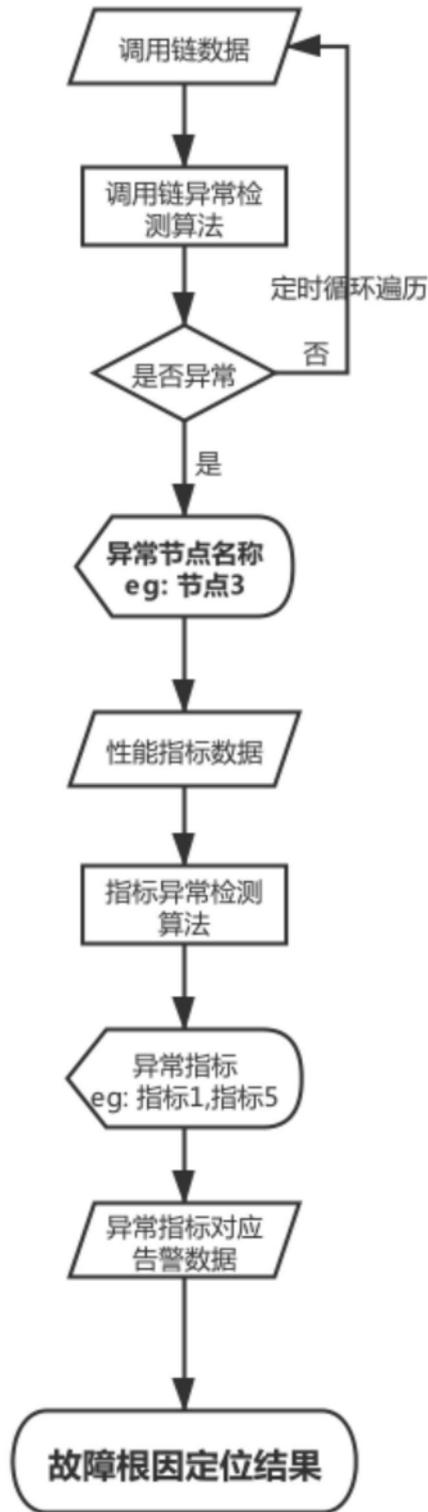


图6

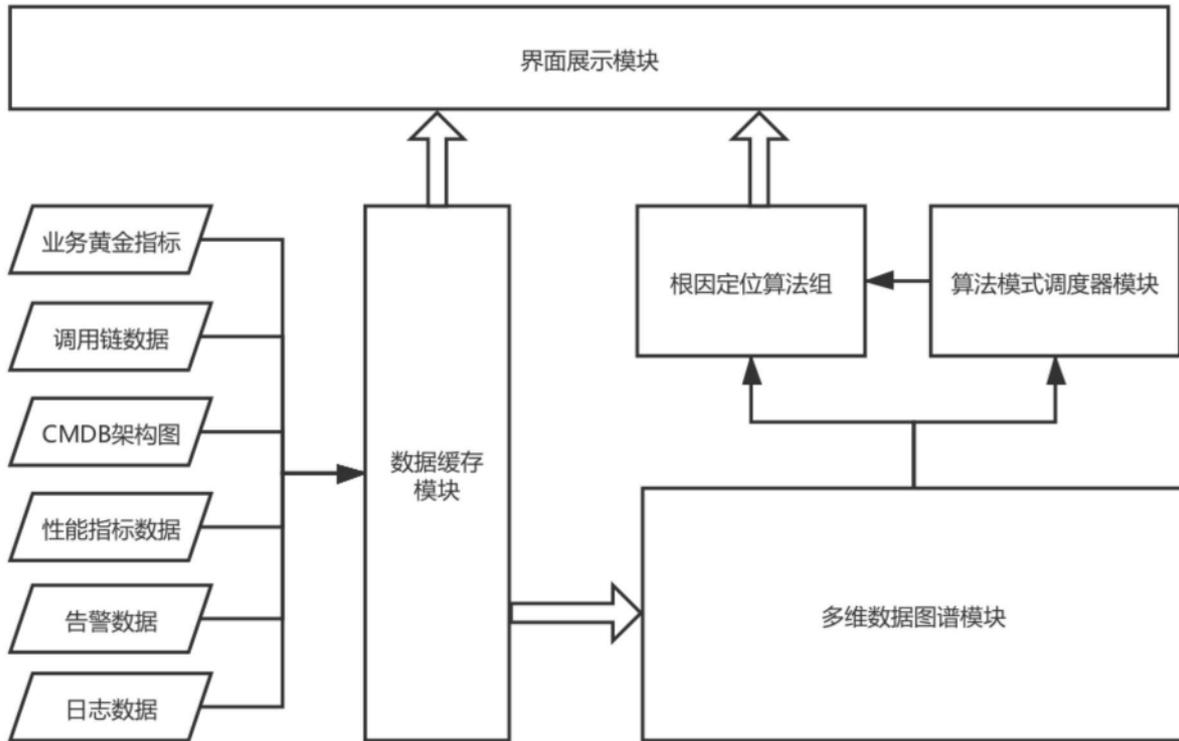


图7