



(21) 申请号 202210244751.2

(22) 申请日 2022.03.14

(65) 同一申请的已公布的文献号

申请公布号 CN 114637786 A

(43) 申请公布日 2022.06.17

(73) 专利权人 徐工汉云技术股份有限公司

地址 221122 江苏省徐州市徐州经济技术
开发区科技路6号

(72) 发明人 郭佳宇 胡旭辉 王飞 邢斌

(74) 专利代理机构 南京纵横知识产权代理有限
公司 32224

专利代理师 俞翠华

(51) Int.Cl.

G06F 16/2458 (2019.01)

G06Q 50/30 (2012.01)

(56) 对比文件

CN 108415408 A, 2018.08.17

CN 114116281 A, 2022.03.01

审查员 夏薇

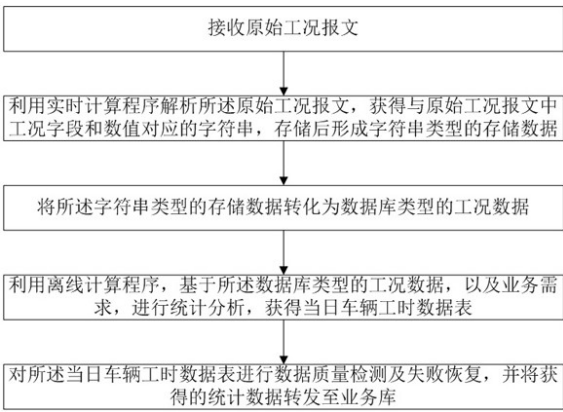
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种车辆工时离线计算方法及存储介质

(57) 摘要

本发明公开了一种车辆工时离线计算方法及存储介质,所述方法包括接收原始工况报文;利用实时计算程序解析所述原始工况报文,获得与原始工况报文中工况字段和数值对应的字符串,存储后形成字符串类型的存储数据;将所述字符串类型的存储数据转化为数据库类型的工况数据;利用离线计算程序,基于所述数据库类型的工况数据,以及业务需求,进行统计分析,获得当日车辆工时数据表;对所述当日车辆工时数据表进行数据质量检测及失败恢复,并将获得的统计数据转发至业务库。本发明能够有效避免因出现缺失或偏差计算后工作时间数据直接导入业务库,导致客户使用错误数据,也能够提高针对缺失偏差计算处理的及时性。



1. 一种车辆工时离线计算方法,其特征在于,包括:

步骤一、接收车辆上传的原始工况报文;

步骤二、利用实时计算程序解析所述原始工况报文,获得与原始工况报文中工况字段和数值对应的字符串,存储后形成字符串类型的存储数据;

步骤三、将所述字符串类型的存储数据转化为数据库类型的工况数据;

步骤四、利用离线计算程序,基于所述数据库类型的工况数据,以及业务需求,进行统计分析,获得当日车辆工时数据表;

步骤五、对所述当日车辆工时数据表进行数据质量检测及失败恢复,并将获得的统计数据转发至业务库;

所述对所述当日车辆工时数据表进行数据质量检测及失败恢复,具体包括:

根据数据规则检测当日车辆工时数据表;

当判断出所述当日车辆工时数据表中的数据不符合要求,则判断工况表中数据量是否正常;所述工况表中的数据是通过对记录不同公司的车辆表按公司分类解析得来;

若数据量正常,说明步骤一至步骤三正常,步骤四异常,则需要重启离线计算程序;若数据量不正常,说明步骤一至步骤三中某一步骤出现异常,则判断车辆上传的原始数据与记录的车辆表中哪一个出现问题,并对出现问题的数据进行恢复;

所述原始工况报文中包括车辆工况信息;所述数据规则检测包括:

当日车辆工时数据表的车辆数量应与记录所有车辆的车辆表一致,若车辆当日没有工作且未上传车辆工况信息,则该车辆当日工时应为累计使用昨日的总工作时间;

若今日发动机累计总工时减去昨日发动机累计总工时大于今日工时,则说明原始数据存在工况数据的跳变,需要按照车辆上传的GPS时间顺序将车辆工况信息排序,如果两条车辆工况信息的发动机累计总工时信息之差大于GPS时间,或者随着时间的增加发动机累计总工时降低,则说明该原始数据为无效跳变数据,需要剔除该无效跳变数据。

2. 根据权利要求1所述的一种车辆工时离线计算方法,其特征在于:所述对所述当日车辆工时数据表进行数据质量检测及失败恢复,还包括:

当判断出所述当日车辆工时数据表中的数据符合要求,则判定为当日数据正常,并发送到业务数据库。

3. 根据权利要求1所述的一种车辆工时离线计算方法,其特征在于:所述判断车辆上传的原始数据与记录的车辆表中哪一个出现问题,并对出现问题的数据进行恢复,具体包括以下步骤:

若判断出车辆上传的原始数据的数据量远少于之前正常日期的数据量,则表示步骤二异常,需要重启离线计算程序;

若判断出车辆上传的原始数据的数据量正常,且车辆表中的车辆数远少于预设值,则说明车辆表有缺失,导致工况缺失,需要重新拉取车辆表后重启离线计算程序。

4. 根据权利要求1所述的一种车辆工时离线计算方法,其特征在于:所述利用实时计算程序解析所述原始工况报文,获得与原始工况报文中工况字段和数值对应的字符串,存储后形成字符串类型的存储数据,具体为:

终端会根据设定每隔一段时间上传包含车辆工况信息的原始报文,原始报文会先被发送到网关,经网关确认终端为终端表中有记录的终端之后;

由实时计算程序根据车辆对应通信协议与功能集将原始报文转换为工况字段和数值对应的字符串,并存入HDFS中,形成字符串类型的存储数据。

5.根据权利要求1所述的一种车辆工时离线计算方法,其特征在于:将所述字符串类型的存储数据转化为数据库类型的工况数据,具体为:

利用数据仓库抽取、转化、加载、计算所述字符串类型的存储数据,将其转化为数据库类型的工况数据。

6.一种存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现权利要求1-5任一项所述方法的步骤。

一种车辆工时离线计算方法及存储介质

技术领域

[0001] 本发明属于大数据离线计算领域，具体涉及一种车辆工时离线计算方法及存储介质。

背景技术

[0002] 目前，大数据离线计算由于其数据量大、保存时间长、方便查询等优点，已广泛应用于数据分析、数据挖掘等领域，已经有很多涉及离线计算数据获取、流程调度、结果推送、数据检测等过程的方案产品，其中数据检测包括数据存储到数据库之后的检测以及入库前实时分析筛选数据。

[0003] 目前涉及离线计算质量检测的方案绝大多数为设置筛选过滤条件后自动选取所需数据后存入数据仓库方便后续计算，或者在离线数据为成功发送到目标数据库后重传，存在因为出现缺失或偏差计算后工作时间数据直接导入业务库，导致客户使用错误数据。

发明内容

[0004] 针对上述问题，本发明提出一种车辆工时离线计算方法及存储介质，能够有效避免因出现缺失或偏差计算后工作时间数据直接导入业务库，导致客户使用错误数据，也能够提高针对缺失偏差计算处理的及时性。

[0005] 为了实现上述技术目的，达到上述技术效果，本发明通过以下技术方案实现：

[0006] 第一方面，本发明提供了一种车辆工时离线计算方法，包括：

[0007] 步骤一、接收车辆上传的原始工况报文；

[0008] 步骤二、利用实时计算程序解析所述原始工况报文，获得与原始工况报文中工况字段和数值对应的字符串，存储后形成字符串类型的存储数据；

[0009] 步骤三、将所述字符串类型的存储数据转化为数据库类型的工况数据；

[0010] 步骤四、利用离线计算程序，基于所述数据库类型的工况数据，以及业务需求，进行统计分析，获得当日车辆工时数据表；

[0011] 步骤五、对所述当日车辆工时数据表进行数据质量检测及失败恢复，并将获得的统计数据转发至业务库。

[0012] 可选地，所述对所述当日车辆工时数据表进行数据质量检测及失败恢复，具体包括：

[0013] 根据数据规则检测当日车辆工时数据表；

[0014] 当判断出所述当日车辆工时数据表中的数据不符合要求，则判断工况表中数据量是否正常；所述工况表中的数据是通过记录不同公司的车辆表按公司分类解析得来；

[0015] 若数据量正常，说明步骤一至步骤三正常，步骤四异常，则需要重启离线计算程序；

[0016] 若数据量不正常，说明步骤一至步骤三中某一步骤出现异常，则判断车辆上传的原始数据与记录的车辆表中哪一个出现问题，并对出现问题的数据进行恢复。

[0017] 可选地,所述原始工况报文中包括车辆工况信息;所述数据规则检测包括:

[0018] 当日车辆工时数据表的车辆数量应与记录所有车辆的车辆表一致,若车辆当日没有工作且未上传车辆工况信息,则该车辆当日工时应为累计使用昨日的总工作时间;

[0019] 若今日发动机累计总工时减去昨日发动机累计总工时大于今日工时,则说明原始数据存在工况数据的跳变,需要按照车辆上传的GPS时间顺序将车辆工况信息排序,如果两条车辆工况信息的发动机累计总工时信息之差大于GPS时间,或者随着时间的增加发动机累计总工时降低,则说明该条数据为无效跳变数据,需要剔除该无效跳变数据。

[0020] 可选地,所述对所述当日车辆工时数据表进行数据质量检测及失败恢复,还包括:

[0021] 当判断出所述当日车辆工时数据表中的数据符合要求,则判定为当日数据正常,并发送到业务数据库。

[0022] 可选地,所述判断车辆上传的原始数据与记录的车辆表中哪一个出现问题,并对出现问题的数据进行恢复,具体包括以下步骤:

[0023] 若判断出车辆上传的原始数据的数据量远少于之前正常日期的数据量,则表示步骤二异常,需要重启离线计算程序;

[0024] 若判断出车辆上传的原始数据的数据量正常,且车辆表中的车辆数远少于预设值,则说明车辆表有缺失,导致工况缺失,需要重新拉取车辆表后重启离线计算程序。

[0025] 可选地,所述利用实时计算程序解析所述原始工况报文,获得与原始工况报文中工况字段和数值对应的字符串,存储后形成字符串类型的存储数据,具体为:

[0026] 终端会根据设定每隔一段时间上传包含车辆工况信息的原始报文,原始报文会先被发送到网关,经网关确认终端为终端表中有记录的终端之后;

[0027] 由实时计算程序根据车辆对应通信协议与功能集将原始报文转换为工况字段和数值对应的字符串,并存入HDFS中,形成字符串类型的存储数据。

[0028] 可选地,所述将字符串类型的存储数据转化为数据库类型的工况数据具体为:

[0029] 利用数据仓库抽取、转化、加载、计算所述字符串类型的存储数据,将其转化为数据库类型的工况数据。

[0030] 第二方面,本发明提供了一种存储介质,其上存储有计算机程序,该程序被处理器执行时实现第一方面任一项所述方法的步骤。

[0031] 与现有技术相比,本发明的有益效果:

[0032] 本发明能够在每日工作时间离线计算后自动针对挖掘机工作时间离线计算数据进行检测,如果当日工作时间数据量正常且符合业务需求,则按时导入事业部数据库方便车主进行查询比对。当数据出现数据量缺失、因出现错误上传数据导致计算结果偏差错误等情况时,会自动识别诊断故障原因进行结果分析,并针对故障的原因进行数据恢复重算,在流程结束之后继续进行诊断,确认无误后导入业务库,能够有效避免因为出现缺失或偏差计算后工作时间数据直接导入业务库,导致客户使用错误数据,也能够提高针对缺失偏差计算处理的及时性。

附图说明

[0033] 为了使本发明的内容更容易被清楚地理解,下面根据具体实施例并结合附图,对本发明作进一步详细的说明,其中:

[0034] 图1为本发明一种实施例的车辆工时离线计算方法的流程示意图；

[0035] 图2为本发明一种实施例的数据质量检测及失败恢复的流程示意图。

具体实施方式

[0036] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明的保护范围。

[0037] 下面结合附图对本发明的应用原理作详细的描述。

[0038] 实施例1

[0039] 本发明实施例中一种车辆工时离线计算方法，如图1所示，包括以下步骤：

[0040] 步骤一、接收原始工况报文；所述原始工况报文包括终端（即车辆）根据通信协议定时上传的车辆工况信息，所述车辆工况信息包括如行驶轨迹、经纬度、时间、油耗、里程、总工作时间等信息，以便统计相应数据；

[0041] 步骤二、利用实时计算程序解析所述原始工况报文，获得与原始工况报文中工况字段和数值对应的字符串，存储至HDFS后形成字符串类型的存储数据；

[0042] 步骤三、将所述字符串类型的存储数据转化为数据库类型的工况数据；在具体实施过程中，所述步骤三具体为利用数据仓库抽取、转化、加载、计算所述字符串类型的存储数据，将其转化为数据库类型的工况数据；

[0043] 步骤四、利用离线计算程序，基于所述数据库类型的工况数据，以及业务需求，进行统计分析，获得当日车辆工时数据表；

[0044] 步骤五、对所述当日车辆工时数据表进行数据质量检测及失败恢复，并将获得的统计数据转发至业务库。

[0045] 在本发明实施例的一种具体实施方式中，所述利用实时计算程序解析所述原始工况报文，获得与原始工况报文中工况字段和数值对应的字符串，存储后形成字符串类型的存储数据，具体为：

[0046] 终端会根据设定每隔一段时间上传包含车辆工况信息的原始报文，原始报文会先被发送到网关，经网关确认终端为终端表中有记录的终端之后；

[0047] 由实时计算程序根据车辆对应通信协议与功能集将原始报文转换为工况字段和数值对应的字符串，并存入HDFS中，形成字符串类型的存储数据。

[0048] 在本发明实施例的一种具体实施方式中，如图2所示，所述对所述当日车辆工时数据表进行数据质量检测及失败恢复，具体包括：

[0049] 根据数据规则检测当日车辆工时数据表；

[0050] 当判断出所述当日车辆工时数据表中的数据符合要求，则判定为当日数据正常，并发送到业务数据库；

[0051] 当判断出所述当日车辆工时数据表中的数据不符合要求，则判断工况表中数据量是否正常；所述工况表的数据是通过记录不同公司的车辆表按公司分类解析得来；

[0052] 若数据量正常，说明步骤一至步骤三正常，步骤四异常，则需要重启离线计算程序；

[0053] 若数据量不正常，说明步骤一至步骤三中某一步骤出现异常，则判断车辆上传的

原始数据与记录的车辆表中哪一个出现问题,并对出现问题的数据进行恢复。

[0054] 在本发明实施例的一种具体实施方式中,所述数据规则具体包括以下:

[0055] 当日车辆工时数据表的车辆数量应与记录所有车辆的车辆表一致,若车辆当日没有工作且未上传车辆工况信息,则该车辆当日工时应为累计使用昨日的总工作时间;

[0056] 若今日发动机累计总工时减去昨日发动机累计总工时大于今日工时,则说明原始数据存在工况数据的跳变,需要按照车辆上传的GPS时间顺序将车辆工况信息排序,如果两条车辆工况信息的发动机累计总工时信息之差大于GPS时间,或者随着时间的增加发动机累计总工时降低,则说明该条数据为无效跳变数据,需要剔除该无效跳变数据。

[0057] 在本发明实施例的一种具体实施方式中,所述判断车辆上传的原始数据与记录的车辆表中哪一个出现问题,并对出现问题的数据进行恢复,具体包括以下步骤:

[0058] 若判断出车辆上传的原始数据的数据量远少于之前正常日期的数据量,则表示步骤二异常,需要重启离线计算程序;

[0059] 若判断出车辆上传的原始数据的数据量正常,且车辆表中的车辆数远少于预设值,则说明车辆表有缺失,导致工况缺失,需要重新拉取车辆表后重启离线计算程序。

[0060] 下面结合一具体实施方式对本发明实施例中的方法的具体流程进行详细说明。

[0061] S10 :终端上传原始工况报文;

[0062] 车辆终端根据通信协议定时上传车辆工况信息,所述车辆工况信息包括如行驶轨迹、经纬度、时间、油耗、里程、总工作时间等信息,以便统计相应数据。

[0063] S20 :实时计算程序解析原始工况报文,并存入HDFS。

[0064] 车辆终端根据设定每隔一段时间上传包含车辆工况信息的原始报文,上传的原始报文先发送到网关,经网关确认终端为终端表中有记录的终端之后,由实时计算程序根据车辆对应通信协议与功能集将原始报文转换为工况字段和数值对应的字符串并存入HDFS中,形成字符串类型的存储数据。

[0065] S30 :数据仓库抽取转化加载计算所需工况字段;

[0066] 字符串类型的存储数据存储在HDFS之后,会经过抽取、转化、加载的过程,过滤掉重复信息和无效工况,并将字符串类型的存储数据转化为数据库类型的工况数据,以便后续计算。

[0067] S40 :每日离线计算统计数据;

[0068] 车辆终端根据需要统计的维度,每天将当日车辆上传的车辆工况信息根据既定的逻辑业务需求进行统计分析,最终得出车辆当日的工时统计信息。

[0069] S50 :统计数据转发业务库数据质量检测及失败恢复。

[0070] 每天计算完成之后需要对数据结果进行检测,符合正常计算的结果之后导入业务库,否则需对错误进行溯源,在失败的步骤处进行恢复并在计算完成后导入业务库;

[0071] S60:统计数据转发业务库。

[0072] 计算完成之后,将每日数据转发到挖掘机业务数据库,方便页面统计展示及客户查询。

[0073] 其中,步骤S50具体包括以下子步骤:

[0074] 步骤S501:根据数据规则检测当日挖掘机车辆工时数据表;数据需要符合以下要求:

[0075] 当日车辆工时数据表中的车辆数量应该和记录所有车辆的车辆表一致,为方便客户查询使用,即使车辆当日没有工作且未上传车辆工况信息,也应该将车辆当日工作时间置零并累计使用昨日的总工作时间。

[0076] 车辆表当日工时与前一日发动机累计总工时之和等于当日发动机累计总工时,以确保数据的连续性,如果今日发动机累计总工时减去昨日发动机累计总工时大于今日工作时间,则说明原始数据存在工况数据的跳变,需要针对工况数据进行进一步的比对分析,按照车辆上传的GPS时间顺序将车辆工况信息排序,如果两条工况的发动机累计总工时信息之差大于GPS时间,或者随着时间的增加发动机累计总工时降低,则说明该条数据为无效跳变数据,需要剔除跳变数据对总工时的影响。

[0077] 本步骤中的数据检测针对挖掘机车辆的具体工作情况和客户使用习惯,有针对性的使用了数据检验方法。

[0078] 步骤S502:根据步骤S501判断当日车辆表数据是否符合要求,如果是,判断为当日数据正常,可以发送到业务数据库;如果否,转步骤步骤S503。

[0079] 步骤S503:每日车辆工作时间数据由包含车辆当日上传所有工况数据的工况表计算而来,因此在当日离线计算的工作时间结果不符合要求时,需先判断数据来源是否有误,即判断工况表中数据量是否正常。查询当日工况与前几日工况的对比,若数据量正常,则说明步骤一至步骤三正常,而步骤四的离线计算任务步骤出现异常失败,需要重启离线计算程序。若工况表数据量明显少于前几日,则说明步骤一至步骤三中某一步骤出现错误,转步骤S504进行下一步判断。

[0080] 步骤S504:若工况表中的数据量异常,则需要判断车辆上传的原始数据和记录公司分类的车辆表中哪一个出现问题;其中,工况表中的数据由记录不同公司的车辆表按公司分类解析得来。

[0081] 首先对车辆上传的原始数据数据量进行判断,如果数据量远少于之前正常日期的数据量,则是HDFS数据搬运出错,需要重新搬运当日数据后重启离线计算流程。

[0082] 如果原始数据量正常,且车辆表车辆远少于预设值,则说明车辆表缺失车辆导致工况缺失,需要从事业部平台重新拉取车辆表数据后重启离线计算流程。

[0083] 实施例2

[0084] 基于与实施例1相同的发明构思,本发明实施例中提供了一种存储介质,其上存储有计算机程序,该程序被处理器执行时实现实施例1中任一项所述方法的步骤。

[0085] 以上显示和描述了本发明的基本原理和主要特征和本发明的优点。本行业的技术人员应该了解,本发明不受上述实施例的限制,上述实施例和说明书中描述的只是说明本发明的原理,在不脱离本发明精神和范围的前提下,本发明还会有各种变化和改进,这些变化和改进都落入要求保护的本发明范围内。本发明要求保护范围由所附的权利要求书及其等效物界定。

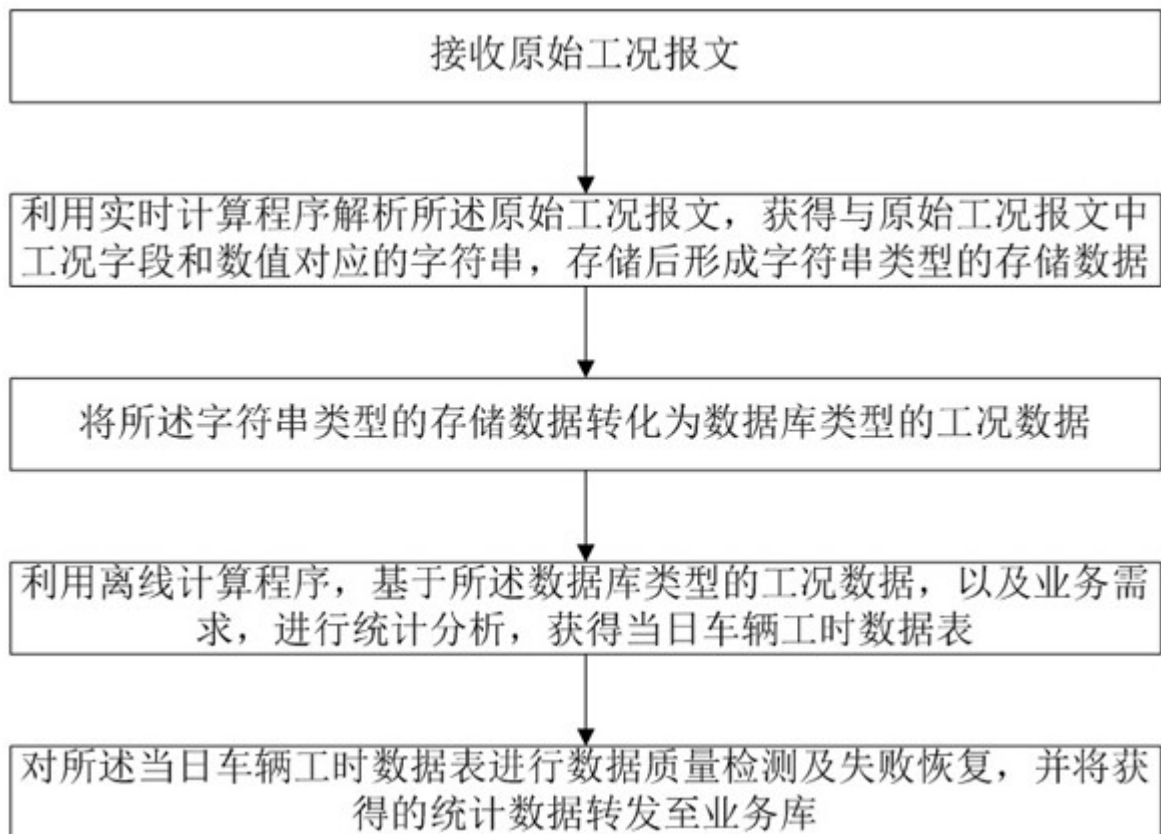


图1

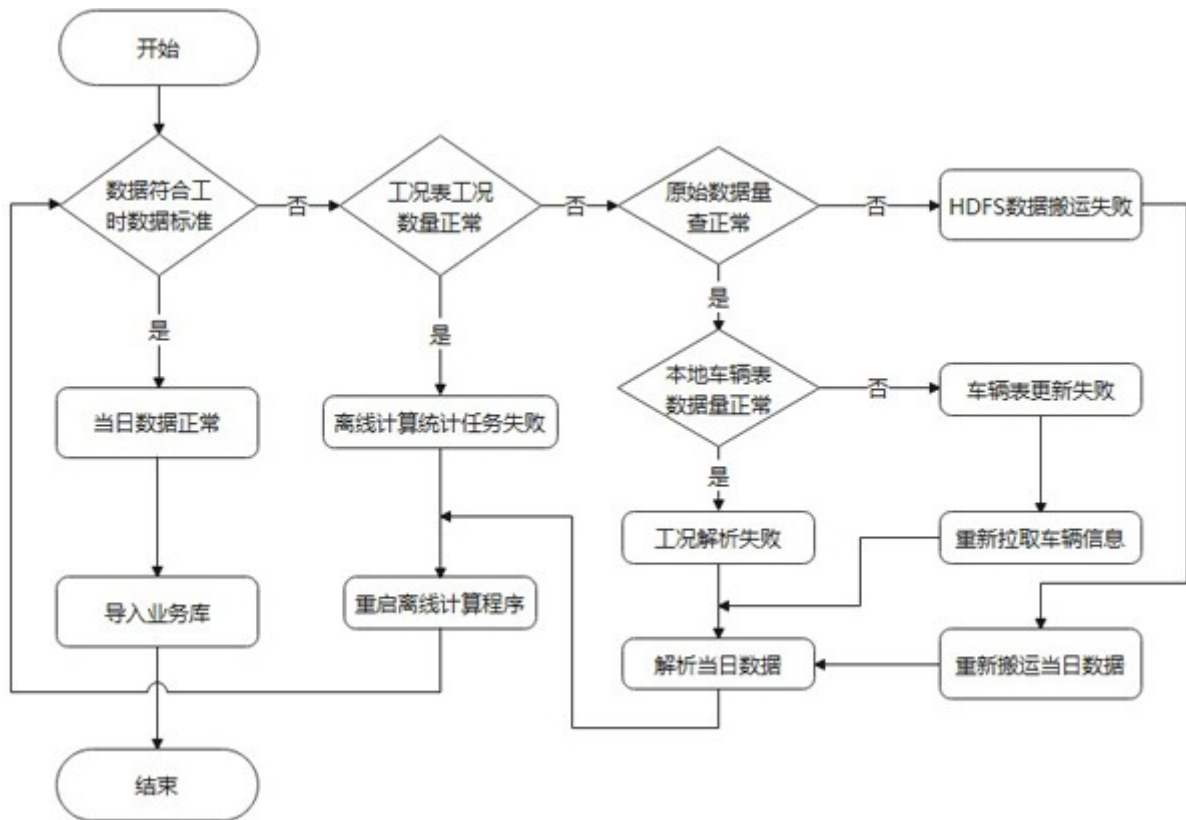


图2