



(12) 发明专利

(10) 授权公告号 CN 103246574 B

(45) 授权公告日 2015. 11. 11

(21) 申请号 201210031020. 6

(22) 申请日 2012. 02. 10

(73) 专利权人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四层 847 号邮箱

(72) 发明人 邓浩

(74) 专利代理机构 北京康信知识产权代理有限责任公司 11240

代理人 吴贵明 余刚

(51) Int. Cl.

G06F 11/00(2006. 01)

(56) 对比文件

CN 101178675 A, 2008. 05. 14,

CN 101364291 A, 2009. 02. 11,

CN 101446914 A, 2009. 06. 03,

US 2007038894 A1, 2007. 02. 15,

CN 101572696 A, 2009. 11. 04,

CN 102073912 A, 2011. 05. 25,

US 6141767 A, 2000. 10. 31,

CN 101021890 A, 2007. 08. 22,

张荣顺. 数据处理系统中确认输入数据

准确性可靠性的几种方法. 《山东建材学院学报》. 1987, 第 1 卷 (第 3 期),

审查员 朱雪珍

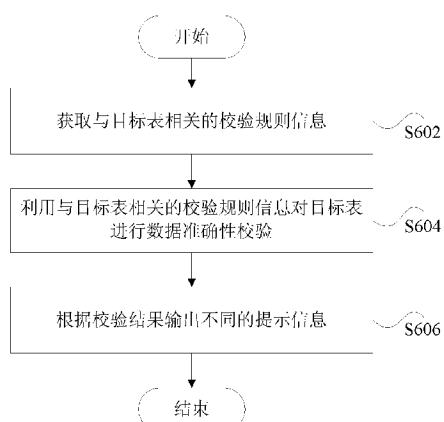
权利要求书3页 说明书10页 附图4页

(54) 发明名称

数据准确性的校验方法及装置

(57) 摘要

本申请公开了一种数据准确性的校验方法及装置。其中，该方法包括：获取与目标表相关的校验规则信息，其中，目标表为需要进行数据准确性校验的目标数据表；利用与目标表相关的校验规则信息对目标表进行数据准确性校验；以及根据校验结果输出不同的提示信息。通过本申请，能够使得对数据准确性的监控校验更及时。



1. 一种数据准确性的校验方法,其特征在于包括 :

获取与目标表相关的校验规则信息,其中,所述目标表为需要进行数据准确性校验的目标数据表;

利用与所述目标表相关的校验规则信息对所述目标表进行数据准确性校验;以及根据校验结果输出不同的提示信息,

其中,利用与所述目标表相关的校验规则信息对所述目标表进行数据准确性校验包括:

根据所述校验规则信息生成目标数据指标值查询程序;

根据所述目标数据指标值查询程序对所述目标表进行查询,得到目标数据指标值;

根据所述校验规则信息生成基准数据指标值查询程序;

根据所述基准数据指标值查询程序对所述目标表进行查询,得到基准数据指标值;

对所述目标数据指标值与所述基准数据指标值进行比较以对所述目标表进行数据准确性校验。

2. 根据权利要求 1 所述的数据准确性的校验方法,其特征在于,获取与目标表相关的校验规则信息包括:

在校验规则配置表中获取与第一目标表相关的校验规则信息;以及

在所述校验规则配置表中获取与第二目标表相关的校验规则信息,

其中,在所述校验规则配置表中配置有与多个目标表相关的校验规则信息。

3. 根据权利要求 2 所述的数据准确性的校验方法,其特征在于,所述方法还包括:

接收用户输入的校验规则信息;以及

根据接收的校验规则信息对所述校验规则配置表中的校验规则信息进行更新。

4. 根据权利要求 1 所述的数据准确性的校验方法,其特征在于,在利用所述目标表相关的校验规则信息对所述目标表进行数据准确性校验之后,所述方法还包括:

将校验结果记录在校验日志表中。

5. 根据权利要求 1 所述的数据准确性的校验方法,其特征在于,

根据校验结果输出不同的提示信息包括:

判断所述目标数据指标值与所述基准数据指标值的差值是否大于第一预定阈值;

在所述目标数据指标值与所述基准数据指标值的差值大于等于所述第一预定阈值时,返回校验失败信息并输出第一告警信息;

在所述目标数据指标值与所述基准数据指标值的差值小于所述第一预定阈值时,判断所述目标数据指标值与所述基准数据指标值的差值是否大于等于第二预定阈值,其中,所述第二预定阈值小于所述第一预定阈值;

在所述目标数据指标值与所述基准数据指标值的差值小于所述第二预定阈值时,返回校验成功信息;以及

在所述目标数据指标值与所述基准数据指标值的差值大于等于所述第二预定阈值时,返回校验成功信息并输出第二告警信息。

6. 根据权利要求 1 所述的数据准确性的校验方法,其特征在于,利用与所述目标表相关的校验规则信息对所述目标表进行数据准确性校验包括采用以下校验规则对所述目标表进行数据准确性校验:

对所述目标表中的数据进行横向校验；

对所述目标表中的数据进行纵向校验；以及

基于预定的指标值对所述目标表中的数据进行校验。

7. 一种数据准确性的校验装置，其特征在于包括：

获取模块，用于获取与目标表相关的校验规则信息，其中，所述目标表为需要进行数据准确性校验的目标数据表；

校验模块，用于利用与所述目标表相关的校验规则信息对所述目标表进行数据准确性校验；以及

提示模块，用于根据校验结果输出不同的提示信息，

其中，所述校验模块包括：

第一生成模块，用于根据所述校验规则信息生成目标数据指标值查询程序；

第一查询模块，用于根据所述目标数据指标值查询程序对所述目标表进行查询，得到目标数据指标值；

第二生成模块，用于根据所述校验规则信息生成基准数据指标值查询程序；

第二查询模块，根据所述基准数据指标值查询程序对所述目标表进行查询，得到基准数据指标值；

校验子模块，用于对所述目标数据指标值与所述基准数据指标值进行比较以对所述目标表进行数据准确性校验。

8. 根据权利要求 7 所述的数据准确性的校验装置，其特征在于，所述获取模块包括：

第一获取模块，用于在校验规则配置表中获取与第一目标表相关的校验规则信息；以及

第二获取模块，用于在所述校验规则配置表中获取与第二目标表相关的校验规则信息，

其中，在所述校验规则配置表中配置有与多个目标表相关的校验规则信息。

9. 根据权利要求 8 所述的数据准确性的校验装置，其特征在于，所述获取模块包括：

接收模块，用于接收用户输入的校验规则信息；以及

更新模块，用于根据接收的校验规则信息对所述校验规则配置表中的校验规则信息进行更新。

10. 根据权利要求 7 所述的数据准确性的校验装置，其特征在于，在利用所述目标表相关的校验规则信息对所述目标表进行数据准确性校验之后，所述装置还包括：

记录模块，用于将校验结果记录在校验日志表中。

11. 根据权利要求 7 所述的数据准确性的校验装置，其特征在于，

所述提示模块包括：

第一判断模块，用于判断所述目标数据指标值与所述基准数据指标值的差值是否大于等于第一预定阈值；

第一提示模块，用于在所述目标数据指标值与所述基准数据指标值的差值大于等于所述第一预定阈值时，返回校验失败信息并输出第一告警信息；

第二判断模块，用于在所述目标数据指标值与所述基准数据指标值的差值小于所述第一预定阈值时，判断所述目标数据指标值与所述基准数据指标值的差值是否大于第二预定

阈值，其中，所述第二预定阈值小于所述第一预定阈值；

第二提示模块，用于在所述目标数据指标值与所述基准数据指标值的差值小于所述第二预定阈值时，返回校验成功信息；以及

第三提示模块，用于在所述目标数据指标值与所述基准数据指标值的差值大于等于所述第二预定阈值时，返回校验成功信息并输出第二告警信息。

12. 根据权利要求 7 所述的数据准确性的校验装置，其特征在于，利用与所述目标表相关的校验规则信息对所述目标表进行数据准确性校验包括采用以下校验规则对所述目标表进行数据准确性校验：

第一校验模块，用于对所述目标表中的数据进行横向校验；

第二校验模块，用于对所述目标表中的数据进行纵向校验；以及

第三校验模块，用于基于预定的指标值对所述目标表中的数据进行校验。

## 数据准确性的校验方法及装置

### 技术领域

[0001] 本申请涉及数据处理领域,具体而言,涉及一种数据准确性的校验方法及装置。

### 背景技术

[0002] 随着数据仓库规模的扩大,应用的不断增多,数据间的依赖关系也越来越复杂。一个统计报表 / 绩效指标,都可能需要来源于多个数据表,并经过多层转换,才能计算得到。同时,数据在企业的日常运营中发挥的作用越来越大,而业务部门应用数据时对数据的首要要求是准确,不准确的数据非但不能辅导运营,提升业务效率,还有可能会误导业务部门,以致做出错误决策。

[0003] 因为数据计算的复杂性,数据出错的机率相对比较大,而数据的使用方不会接受计算错误的数据,因而对数据计算结果进行监控,校验,确保提供正确的结果是非常重要的。

[0004] 目前,一般的数据仓库系统会编写一些数据校验脚本,对计算得到的数据进行校验,这些脚本以数据监控任务方式单独放到调度系统中运行,数据监控任务一般不影响或干涉数据提取、转换和加载 (Extraction-Transformation-Loading, 简称为 ETL) 工作流的运行。当校验得出计算结果有异常时,后续再由工程师查找原因,进行数据的回滚计算操作。

[0005] 对于上述的数据准确性(或者,称作数据质量)的校验方法,通常存在以下几个问题:

[0006] 首先,由于 ETL 任务与数据校验任务分离,数据校验的结果只是作为事后进行故障排查分析的参考,并不能有效地隔离某项数据错误给整个系统带来的影响,因而造成监控处理不够及时,事后的处理成本较高。

[0007] 其次,由于进行数据校验的业务逻辑脚本分散的封装在各个数据校验任务脚本中,重复编写,维护不便,进而造成同类数据校验程序并存,没有通用性。

[0008] 此外,现有的数据准确性校验规则较简单,通常只考虑与源表对比行数,对基于商业指标的监控,考虑较少。

[0009] 针对现有技术中的数据准确性的校验方法往往造成监控处理不够及时的问题,目前尚未提出有效的解决方案。

### 发明内容

[0010] 本申请的主要目的在于提供一种数据准确性的校验方法及装置,以解决现有技术中的数据准确性的校验方法往往造成监控处理不够及时的问题。

[0011] 为了实现上述目的,根据本申请的一个方面,提供了一种数据准确性的校验方法。

[0012] 根据本申请的数据准确性的校验方法包括:获取与目标表相关的校验规则信息,其中,目标表为需要进行数据准确性校验的目标数据表;利用与目标表相关的校验规则信息对目标表进行数据准确性校验;以及根据校验结果输出不同的提示信息。

[0013] 进一步地,获取与目标表相关的校验规则信息包括:在校验规则配置表中获取与第一目标表相关的校验规则信息;以及在校验规则配置表中获取与第二目标表相关的校验规则信息,其中,在校验规则配置表中配置有与多个目标表相关的校验规则信息。

[0014] 进一步地,上述方法还包括:接收用户输入的校验规则信息;以及根据接收的校验规则信息对校验规则配置表中的校验规则信息进行更新。

[0015] 进一步地,在利用目标表相关的校验规则信息对目标表进行数据准确性校验之后,方法还包括:将校验结果记录在校验日志表中。

[0016] 进一步地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括:根据校验规则信息生成目标数据指标值查询程序;根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;根据校验规则信息生成基准数据指标值查询程序;根据基准数据指标值查询程序对目标表进行查询,得到基准数据指标值;对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验。

[0017] 进一步地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括:根据校验规则信息生成目标数据指标值查询程序;根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;以及对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验,根据校验结果输出不同的提示信息包括:判断目标数据指标值与基准数据指标值的差值是否大于等于第一预定阈值;在目标数据指标值与基准数据指标值的差值大于等于第一预定阈值时,返回校验失败信息并输出第一告警信息;在目标数据指标值与基准数据指标值的差值小于第一预定阈值时,判断目标数据指标值与基准数据指标值的差值是否大于等于第二预定阈值,其中,第二预定阈值小于第一预定阈值;在目标数据指标值与基准数据指标值的差值小于第二预定阈值时,返回校验成功信息;以及在目标数据指标值与基准数据指标值的差值大于等于第二预定阈值时,返回校验成功信息并输出第二告警信息。

[0018] 进一步地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括采用以下校验规则对目标表进行数据准确性校验:对目标表中的数据进行横向校验;对目标表中的数据进行纵向校验;以及基于预定的指标值对目标表中的数据进行校验。

[0019] 为了实现上述目的,根据本申请的另一方面,提供了一种数据准确性的校验装置。该数据准确性的校验装置包括:获取模块,用于获取与目标表相关的校验规则信息,其中,目标表为需要进行数据准确性校验的目标数据表;校验模块,用于利用与目标表相关的校验规则信息对目标表进行数据准确性校验;以及提示模块,用于根据校验结果输出不同的提示信息。

[0020] 进一步地,获取模块包括:第一获取模块,用于在校验规则配置表中获取与第一目标表相关的校验规则信息;以及第二获取模块,用于在校验规则配置表中获取与第二目标表相关的校验规则信息,其中,在校验规则配置表中配置有与多个目标表相关的校验规则信息。

[0021] 进一步地,获取模块包括:接收模块,用于接收用户输入的校验规则信息;以及更新模块,用于根据接收的校验规则信息对校验规则配置表中的校验规则信息进行更新。

[0022] 进一步地,在利用目标表相关的校验规则信息对目标表进行数据准确性校验之后,方法还包括:记录模块,用于将校验结果记录在校验日志表中。

[0023] 进一步地,校验模块包括:第一生成模块,用于根据校验规则信息生成目标数据指标值查询程序;第一查询模块,用于根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;第二生成模块,用于根据校验规则信息生成基准数据指标值查询程序;第二查询模块,根据基准数据指标值查询程序对目标表进行查询,得到基准数据指标值;校验子模块,用于对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验。

[0024] 进一步地,校验模块包括:第一生成模块,用于根据校验规则信息生成目标数据指标值查询程序;第一查询模块,用于根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;以及校验子模块,用于对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验,提示模块包括:第一判断模块,用于判断目标数据指标值与基准数据指标值的差值是否大于等于第一预定阈值;第一提示模块,用于在目标数据指标值与基准数据指标值的差值大于等于第一预定阈值时,返回校验失败信息并输出第一告警信息;第二判断模块,用于在目标数据指标值与基准数据指标值的差值小于第一预定阈值时,判断目标数据指标值与基准数据指标值的差值是否大于等于第二预定阈值,其中,第二预定阈值小于第一预定阈值;第二提示模块,用于在目标数据指标值与基准数据指标值的差值小于第二预定阈值时,返回校验成功信息;以及第三提示模块,用于在目标数据指标值与基准数据指标值的差值大于等于第二预定阈值时,返回校验成功信息并输出第二告警信息。

[0025] 进一步地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括采用以下校验规则对目标表进行数据准确性校验:第一校验模块,用于对目标表中的数据进行横向校验;第二校验模块,用于对目标表中的数据进行纵向校验;以及第三校验模块,用于基于预定的指标值对目标表中的数据进行校验。

[0026] 通过本申请,采用获取与目标表相关的校验规则信息,其中,目标表为需要进行数据准确性校验的目标数据表;利用与目标表相关的校验规则信息对目标表进行数据准确性校验;以及根据校验结果输出不同的提示信息,由于在进行数据准确性校验之后根据校验结果输出不同的提示信息,因而可以根据提示信息进行及时处理,从而解决了现有技术中的数据准确性的校验方法往往造成监控处理不够及时的问题,进而达到了使得对数据准确性的监控校验更及时的效果。

## 附图说明

[0027] 构成本申请的一部分的附图用来提供对本申请的进一步理解,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

- [0028] 图1是根据本申请第一实施例的数据准确性的校验装置示意图;
- [0029] 图2是根据本申请第二实施例的数据准确性的校验装置示意图;
- [0030] 图3是根据本申请第三实施例的数据准确性的校验装置示意图;
- [0031] 图4是根据本申请第四实施例的数据准确性的校验装置示意图;
- [0032] 图5是根据本申请第五实施例的数据准确性的校验装置示意图;
- [0033] 图6是根据本申请第六实施例的数据准确性的校验装置示意图;
- [0034] 图7是根据本申请实施例的数据准确性的校验系统示意图;

[0035] 图 8 是根据本申请实施例的数据准确性的校验方法流程图 ; 以及

[0036] 图 9 是根据图 7 所示的校验系统的数据准确性的校验方法流程图。

## 具体实施方式

[0037] 需要说明的是, 在不冲突的情况下, 本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0038] 实施例 1

[0039] 图 1 是根据本申请第一实施例的数据准确性的校验装置示意图。

[0040] 如图 1 所示, 该数据准确性的校验装置包括获取模块 10、校验模块 20 和提示模块 30。

[0041] 其中, 获取模块 10, 用于获取与目标表相关的校验规则信息, 其中, 目标表为需要进行数据准确性校验的目标数据表, 其中, 校验规则信息可以是存储在预先配置的表格中, 校验规则是指对数据准确性进行校验的逻辑规则, 是判断数据是否准确的标准。

[0042] 校验模块 20, 用于利用与目标表相关的校验规则信息对目标表进行数据准确性校验, 具体地, 校验模块 20 可以是利用与目标表相关的一条或者多条校验规则信息对目标表进行数据准确性校验, 也可以是利用与目标表相关的校验规则信息对目标表预定范围的数据进行数据准确性校验, 例如, 仅对目标表中的行数据进行准确性校验。

[0043] 提示模块 30, 用于根据不同的校验结果输出不同的提示信息。例如, 在校验成功时, 输出成功信息, 在校验失败时, 输出失败信息, 从而实现对校验结果进行监控的目的。

[0044] 本申请通过定义校验数据准确性的校验模块 20, 为数据仓库 ETL 过程提供通用的数据准确性校验功能, 能对 ETL 过程中生成的数据进行实时的数据准确性校验, 在校验过程中发现的问题, 可以即时通知相关责任人员进行处理。以缩短数据问题发现的时间, 降低事后处理的成本。

[0045] 在该数据准确性的校验装置中, 由于在通过校验模块 20 对目标表进行数据准确性校验之后, 提示模块 30 根据不同的校验结果输出不同的提示信息, 因而可以根据提示信息进行及时的处理, 从而使得对数据准确性的监控校验更及时。

[0046] 本申请实施例中所提供的数据准确性的校验装置可以应用于数据处理系统中, 以供具体的数据处理计算任务, 例如, ETL 任务, 进行数据的准确性校验。其中, 数据处理系统中的 ETL 任务处理单元通过访问数据准确性校验规则信息, 提取相应的准确性校验规则, 然后由校验装置进行数据计算结果的准确性校验, 并且还可以根据校验结果进行错误报警或校验成功等后续操作。

[0047] 实施例 2

[0048] 为了提高校验规则的通用性, 作为实施例 1 的一种优选实施方式, 获取模块 10 可以包括第一获取模块 101 和第二获取模块 102, 如图 2 所示。

[0049] 第一获取模块 101 用于在校验规则配置表中获取与第一目标表相关的校验规则信息; 第二获取模块 102 用于在校验规则配置表中获取与第二目标表相关的校验规则信息, 其中, 在校验规则配置表中配置有与多个目标表相关的校验规则信息。

[0050] 本申请通过在数据仓库中定义一个通用的校验规则配置表, 以目标表为中心, 配置数据校验规则, 进一步地, 通过在数据准确性的校验装置中设置多个获取模块, 多个获取

模块均从该通用的校验规则配置表中获取与不同目标表,例如,第一目标表和第二目标表,对应的校验规则信息以对不同目标表中的数据进行准确性校验,实现了校验规则的统一配置。此外,由于同一校验规则配置表中的校验规则信息能够用来对不同目标表中的数据进行准确性校验,从而提高了校验规则的通用性。

[0051] 实施例 3

[0052] 为了方便对校验规则进行更新,以及便于维护,作为实施例 1 的一种优选方式,获取模块 10 可以包括接收模块 103 和更新模块 104,如图 3 所示。

[0053] 接收模块 103 用于接收用户输入的校验规则信息,更新模块 104 用于根据接收的校验规则信息对校验规则配置表中的校验规则信息进行更新。

[0054] 该实施例中的获取模块 10 可以仅包括接收模块 103 和更新模块 104,也可以同时包括第一获取模块 101、第二获取模块 102、接收模块 103 和更新模块 104。

[0055] 如上述实施例 2 所述,本申请通过通用的校验规则配置表实现了校验规则的统一配置,从而在需要对校验规则配置表中的校验规则进行更新时,只需要通过接收模块 103 输入需要更新的校验规则信息,然后更新模块 104 即可以通过接收模块 103 接收到的信息对校验规则配置表中的校验规则信息进行更新。

[0056] 由于校验规则配置统一,因而维护简单方便,可以对数据校验的规则进行统一配置,所有规则集中在一个数据表中,调整数据校验规则,只需更新相应数据表中的逻辑即可,不需要对 ETL 任务,调度任务,监控任务做出变更,而且其他程序引用更加方便,不再需要单独编写校验程序。

[0057] 优选地,在本申请的任一实施例中,数据准确性的校验装置还可以包括记录模块 40,该记录模块 40 用于在利用目标表相关的校验规则信息对目标表进行数据准确性校验之后,将校验结果记录在校验日志表中,从而便于对错误的数据进行查询。如图 3 所示。

[0058] 实施例 4

[0059] 图 4 是根据本申请第四实施例的数据准确性的校验装置示意图。

[0060] 为了使得数据准确性的校验更加准确,作为上述实施例的一种优选实施方式,校验模块 20 还可以包括:

[0061] 第一生成模块 211,用于根据校验规则信息生成目标数据指标值查询程序;

[0062] 第一查询模块 212,用于根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;

[0063] 第二生成模块 213,用于根据校验规则信息生成基准数据指标值查询程序;

[0064] 第二查询模块 214,根据基准数据指标值查询程序对目标表进行查询,得到基准数据指标值;

[0065] 校验子模块 215,用于对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验。

[0066] 在该实施例中,不只通过第一生成模块 211 和第一查询模块 212 得到目标数据指标值,而且通过设置第二生成模块 213 和第二查询模块 214 来得到基准数据指标值,即,基准数据也是动态的,相比于固定不变的基准数据,能够使得数据准确性的校验更加准确。

[0067] 实施例 5

[0068] 图 5 是根据本申请第四实施例的数据准确性的校验装置示意图。

[0069] 为了使得数据准确性的校验装置的提示更加及时和准确,作为本申请的一种优选实施方式,校验模块 20 包括第一生成模块 221、第一查询模块 222 和校验子模块 223,提示模块 30 包括第一判断模块 301、第一提示模块 302、第二判断模块 303、第二提示模块 304 和第三提示模块 305。

[0070] 第一生成模块 221,用于根据校验规则信息生成目标数据指标值查询程序;第一查询模块 222,用于根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;校验子模块 223,用于对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验,

[0071] 第一判断模块 301,用于判断目标数据指标值与基准数据指标值的差值是否大于等于第一预定阈值;第一提示模块 302,用于在目标数据指标值与基准数据指标值的差值大于等于第一预定阈值时,返回校验失败信息并输出第一告警信息;第二判断模块 303,用于在目标数据指标值与基准数据指标值的差值小于第一预定阈值时,判断目标数据指标值与基准数据指标值的差值是否大于等于第二预定阈值,其中,第二预定阈值小于第一预定阈值;第二提示模块 304,用于在目标数据指标值与基准数据指标值的差值小于等于第二预定阈值时,返回校验成功信息;以及第三提示模块 305,用于在目标数据指标值与基准数据指标值的差值大于第二预定阈值时,返回校验成功信息并输出第二告警信息。

[0072] 在该实施例中,不只在校验成功或者失败时返回相应的信息,而且在校验成功的情况下,利用第二判断模块 303 继续判断是否需要发岀告警信息,其中,在满足提醒阈值的情况下,发出告警信息,从而可以使用户预先获知校验准确性较低的数据,保证了数据的及时性和准确性。

[0073] 实施例 6

[0074] 图 6 是根据本申请第六实施例的数据准确性的校验装置示意图。

[0075] 为了对目标表中的数据进行多角度的校验,作为本申请的一种优选实施方式,校验模块 20 还可以包括第一校验模块 230、第二校验模块 231 和第三校验模块 232,如图 5 所示。

[0076] 第一校验模块 230,用于对目标表中的数据进行横向校验;第二校验模块 231,用于对目标表中的数据进行纵向校验;以及第三校验模块 232,用于基于预定的指标值对目标表中的数据进行校验。其中,横向校验是指将需校验的数据表中的某一汇总指标值,与从其它表计算的同一指标的汇总值进行对比,看差异是否在可接受范围内的方法。纵向校验是指将需校验的数据表中的某一汇总指标值,与以前日期的该表计算的同一指标的汇总值进行对比,看差异是否在可接受范围内的方法。

[0077] 在本实施例中,可以通过多个不同的校验模块从不同的角度对目标表中的数据进行校验,从而使得数据校验更加全面,同时支持的数据校验规则比较丰富,既可以是数据行数的横向,纵向对比,也可以是基于某一商业指标值的对比,对比的数据可以是在同一数据库中,也可以是在不同库中。

[0078] 为了描述的方便,在对本申请实施例的数据准确性的校验装置进行描述时,以功能分为各种单元分别描述。当然,在实施本申请时可以把各单元的功能在同一个或多个软件和 / 或硬件中实现。

[0079] 图 7 是根据本申请实施例的数据准确性的校验系统示意图。

[0080] 如图 7 所示,该校验系统除了包括本申请实施例的数据准确性的校验装置之外,还包括校验规则配置表以及 ETL 任务系统。其中,校验规则配置表与本申请的获取模块 10 相连接,以便获取模块 10 可以从校验规则配置表中获取校验规则信息,ETL 任务系统连接至本申请的校验模块 20,以便 ETL 任务系统可以根据校验结果进行相应的处理。

[0081] ETL 任务是指进行具体数据计算任务的程序,数据计算完成后,可通过调用数据校验执行模块进行数据计算结果的准确性校验,数据校验模块接收到要进行校验的数据表参数后,从数据校验规则配置模块中获取校验规则,进行数据的准确性校验,校验结果表明是数据计算结果错误或存疑时,通过调用报警系统,将错误或警告信息传送给维护人员。

[0082] ETL 任务系统和报警系统是在整个数据仓库系统中的独立子系统,与数据校验功能存在交互,相互调用关系。非本申请所描述的数据校验系统具体实现内容。

[0083] 以下对本申请实施例提供的数据准确性的校验方法进行描述。根据本申请实施例的数据准确性的校验方法可以采用上述本申请实施例中的数据准确性的校验装置的相应模块来执行。

[0084] 图 8 是根据本申请实施例的数据准确性的校验方法流程图。如图 8 所示,该方法包括以下步骤:

[0085] 步骤 S602,获取与目标表相关的校验规则信息,其中,目标表为需要进行数据准确性校验的目标数据表;

[0086] 步骤 S604,利用与目标表相关的校验规则信息对目标表进行数据准确性校验;以及

[0087] 步骤 S606,根据校验结果输出不同的提示信息。

[0088] 在该方法中,由于在对目标表进行数据准确性校验之后,根据不同的校验结果输出不同的提示信息,因而可以根据提示信息进行及时的处理,从而使得对数据准确性的监控校验更及时。

[0089] 优选地,获取与目标表相关的校验规则信息包括:在校验规则配置表中获取与第一目标表相关的校验规则信息;以及在校验规则配置表中获取与第二目标表相关的校验规则信息,其中,在校验规则配置表中配置有与多个目标表相关的校验规则信息。此时,由于同一校验规则配置表中的校验规则信息能够用来对不同目标表中的数据进行准确性校验,从而提高了校验规则的通用性。

[0090] 优选地,上述方法还包括:接收用户输入的校验规则信息;以及根据接收的校验规则信息对校验规则配置表中的校验规则信息进行更新。由于所有规则集中在一个数据表中,调整数据校验规则,只需更新相应数据表中的逻辑即可,不需要对 ETL 任务,调度任务,监控任务做出变更,而且其他程序引用更加方便,不再需要单独编写校验程序。

[0091] 优选地,在利用目标表相关的校验规则信息对目标表进行数据准确性校验之后,上述方法还包括:将校验结果记录在校验日志表中。由于利用目标表相关的校验规则信息对目标表进行数据准确性校验之后,将校验结果记录在校验日志表中,因而便于对错误的数据进行查询。

[0092] 优选地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括:根据校验规则信息生成目标数据指标值查询程序;根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;根据校验规则信息生成基准数据指标值查询程序;根据

基准数据指标值查询程序对目标表进行查询,得到基准数据指标值;对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验。在该实施例中,不只动态得到目标数据指标值,而且动态得到基准数据指标值,相比于固定不变的基准数据,能够使得数据准确性的校验更加准确。

[0093] 优选地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括:根据校验规则信息生成目标数据指标值查询程序;根据目标数据指标值查询程序对目标表进行查询,得到目标数据指标值;以及对目标数据指标值与基准数据指标值进行比较以对目标表进行数据准确性校验。根据校验结果输出不同的提示信息包括:判断目标数据指标值与基准数据指标值的差值是否大于等于第一预定阈值;在目标数据指标值与基准数据指标值的差值大于等于第一预定阈值时,返回校验失败信息并输出第一告警信息;在目标数据指标值与基准数据指标值的差值小于第一预定阈值时,判断目标数据指标值与基准数据指标值的差值是否大于等于第二预定阈值,其中,第二预定阈值小于第一预定阈值;在目标数据指标值与基准数据指标值的差值小于第二预定阈值时,返回校验成功信息;以及在目标数据指标值与基准数据指标值的差值大于等于第二预定阈值时,返回校验成功信息并输出第二告警信息。在该实施例中,不只在校验成功或者失败时返回相应的信息,而且在校验成功的情况下,继续判断是否需要发出告警信息,其中,在满足提醒阈值的情况下,发出告警信息,从而可以使用户预先获知校验准确性较低的数据,保证了数据的及时性和准确性。

[0094] 优选地,利用与目标表相关的校验规则信息对目标表进行数据准确性校验包括采用以下校验规则对目标表进行数据准确性校验:对目标表中的数据进行横向校验;对目标表中的数据进行纵向校验;以及基于预定的指标值对目标表中的数据进行校验。在本实施例中,可以通过多个不同的校验模块从不同的角度对目标表中的数据进行校验,从而使得数据校验更加全面,同时支持的数据校验规则比较丰富,既可以是数据行数的横向,纵向对比,也可以是基于某一商业指标值的对比,对比的数据可以是在同一数据库中,也可以是在不同库中。

[0095] 图 9 是根据图 7 所示的校验系统的数据准确性的校验方法流程图。

[0096] 如图 9 所示,整个方案包括以下一些内容

[0097] 校验规则配置:

[0098] 在数据仓库中定义一个通用的校验规则配置表,以目标表为中心,配置数据校验规则。

[0099] 校验规则调用执行

[0100] 具体地,校验规则执行的流程为:

[0101] 步骤 S1,ETL 任务调用数据校验执行模块,其调用参数为目标表名。

[0102] 步骤 S2,数据校验执行模块从数据校验规则配置存储模块中取得与该表相关的校验规则(可以是任意多个校验规则),进行具体的数据校验。并根据校验结果,进行后续的报警或返回成功状态等动作。

[0103] 具体的调用执行过程为:

[0104] 步骤 S21,根据目标表名参数,从数据校验规则配置表取回所有与目标表相关的校验规则信息,具体地,可以根据需要在校验规则表中设置一下参数:

[0105] 目标数据库名称

- [0106] 目标表名
- [0107] 目标表指标计算规则
- [0108] 目标表数据时间
- [0109] 目标表数据取值时间表达式
- [0110] 源数据库名称
- [0111] 源表名或源子查询
- [0112] 源表指标计算规则
- [0113] 源表数据取值时间表达式
- [0114] 源表数据时间
- [0115] 指标校验错误失败报警阀值
- [0116] 指标校验不一致通知提醒阀值
- [0117] 步骤 S22, 数据校验模块将取到的与该表相关的所有规则, 轮询执行每条规则, 校验规则可以对数据行数或某一商业指标值进行横向, 纵向对比。
- [0118] 步骤 S3, 执行校验规则, 执行一个具体规则的过程可以如下:
- [0119] 步骤 S311, 根据目标数据库名称, 目标表名, 目标表指标计算规则, 目标表数据时间, 目标表数据取值时间表达式等信息, 生成一个计算目标表数据指标值或行数的动态 sql.
- [0120] 步骤 S312, 连上目标数据库, 执行生成的动态 sql, 取得目标数据指标值查询程序。
- [0121] 步骤 S321, 根据源数据库名称, 源表名, 源表指标计算规则, 源表数据时间, 源表数据取值时间表达式等信息, 生成一个计算源表数据指标值或行数的动态 sql。
- [0122] 步骤 S322, 连上源数据库, 执行生成的源动态 sql, 取得源数据指标值。
- [0123] 对该条规则的源表和目标表的指标计算结果进行对比, 规则中配置有失败报警阀值的, 且两者结果差异超过这个阀值, 返回失败, 并调用报警模块, 将错误信息通过手机短信, 邮件发送给相关维护人员。当结果差异小于失败报警阀值, 但规则配置有差异通知提醒阀值的, 将差异信息通过邮件方式发送给相关维护人员, 但不返回出错。
- [0124] ETL 任务接收校验结果, 进行成功或失败处理。
- [0125] 通过以上描述可以看出, 本申请实施例实现了以下技术效果:
- [0126] 监控及时, 错误及警告信息实时通知;
- [0127] 校验规则配置统一, 维护简单方便;
- [0128] 支持的数据校验规则比较丰富。
- [0129] 本说明书中的各个实施例均采用递进的方式描述, 各个实施例之间相同相似的部分互相参见即可, 每个实施例重点说明的都是与其他实施例的不同之处。尤其, 对于系统实施例而言, 由于其基本相似于方法实施例, 所以描述的比较简单, 相关之处参见方法实施例的部分说明即可。
- [0130] 需要说明的是, 在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行, 并且, 虽然在流程图中示出了逻辑顺序, 但是在某些情况下, 可以以不同于此处的顺序执行所示出或描述的步骤。
- [0131] 显然, 本领域的技术人员应该明白, 上述的本申请的各模块或各步骤可以用通用

的计算装置来实现,它们可以集中在单个的计算装置上,或者分布在多个计算装置所组成的网络上,可选地,它们可以用计算装置可执行的程序代码来实现,从而,可以将它们存储在存储装置中由计算装置来执行,或者将它们分别制作成各个集成电路模块,或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样,本申请不限制于任何特定的硬件和软件结合。

[0132] 以上所述仅为本申请的优选实施例而已,并不用于限制本申请,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。



图 1

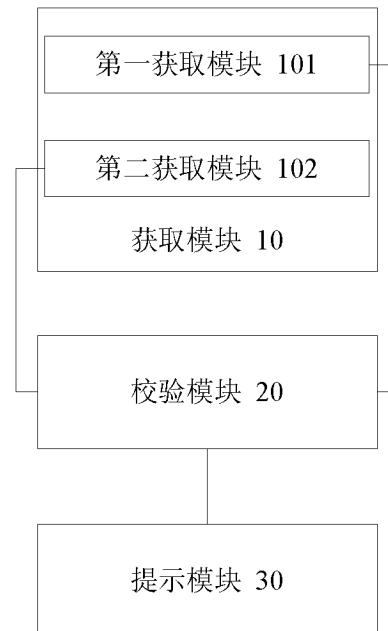


图 2

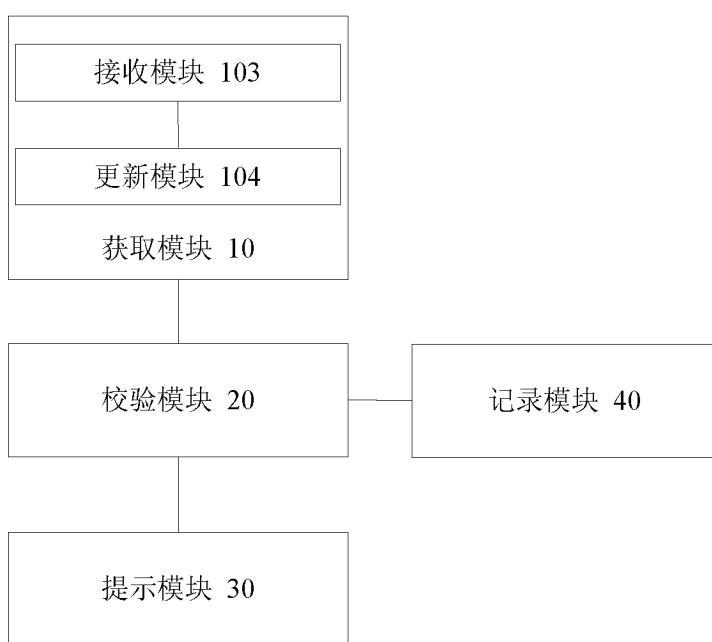


图 3

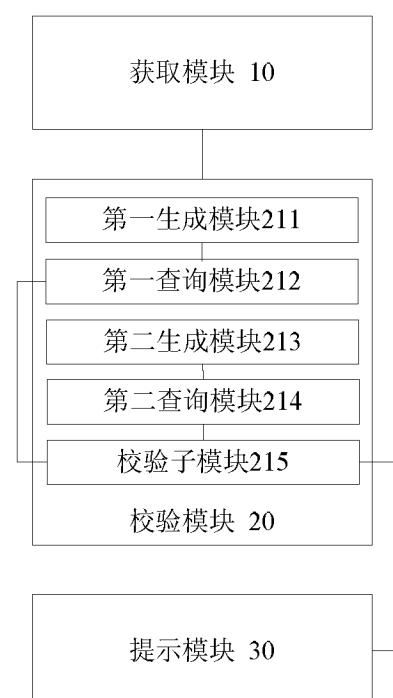


图 4

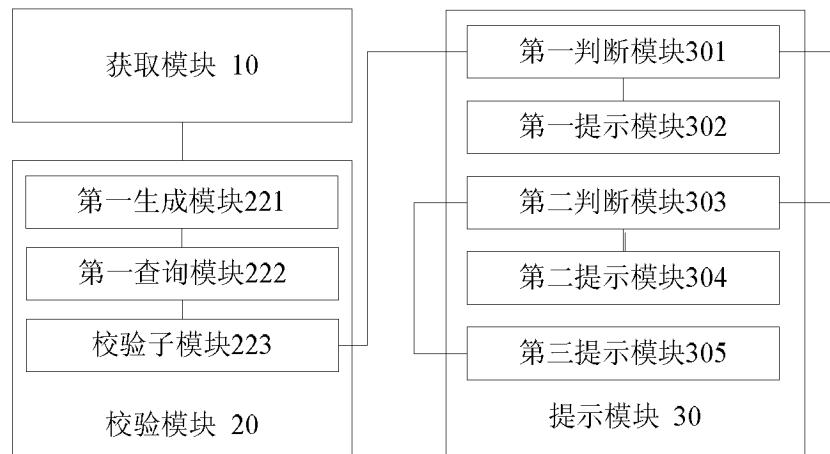


图 5



图 6

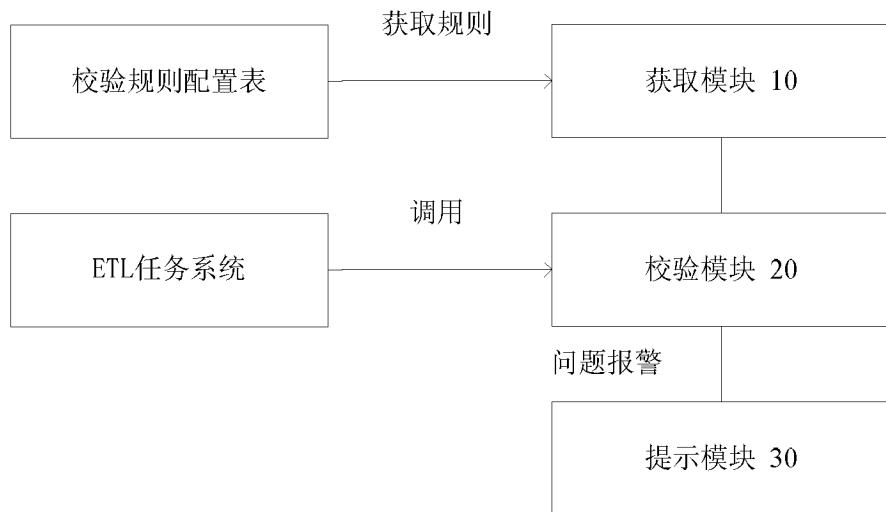


图 7

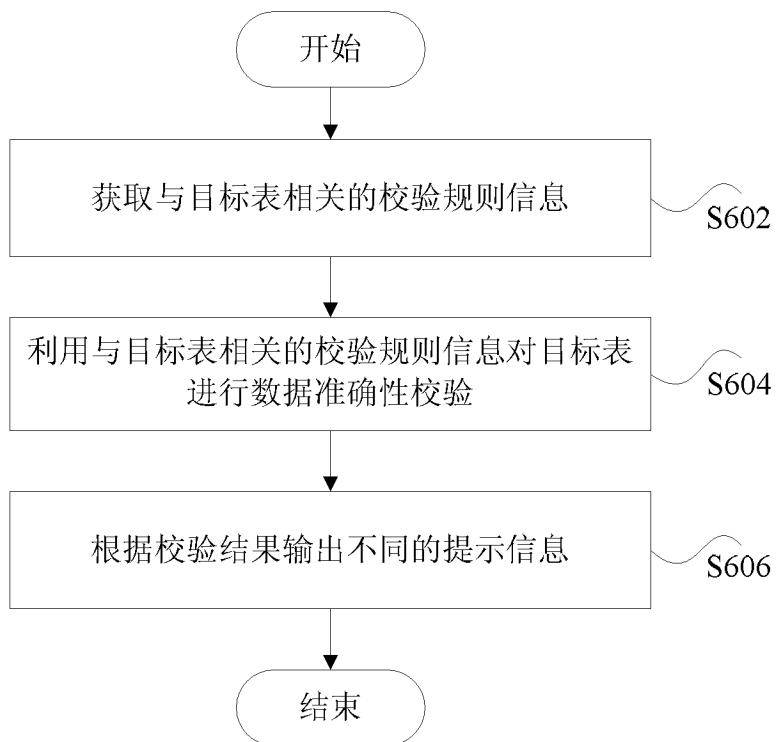


图 8

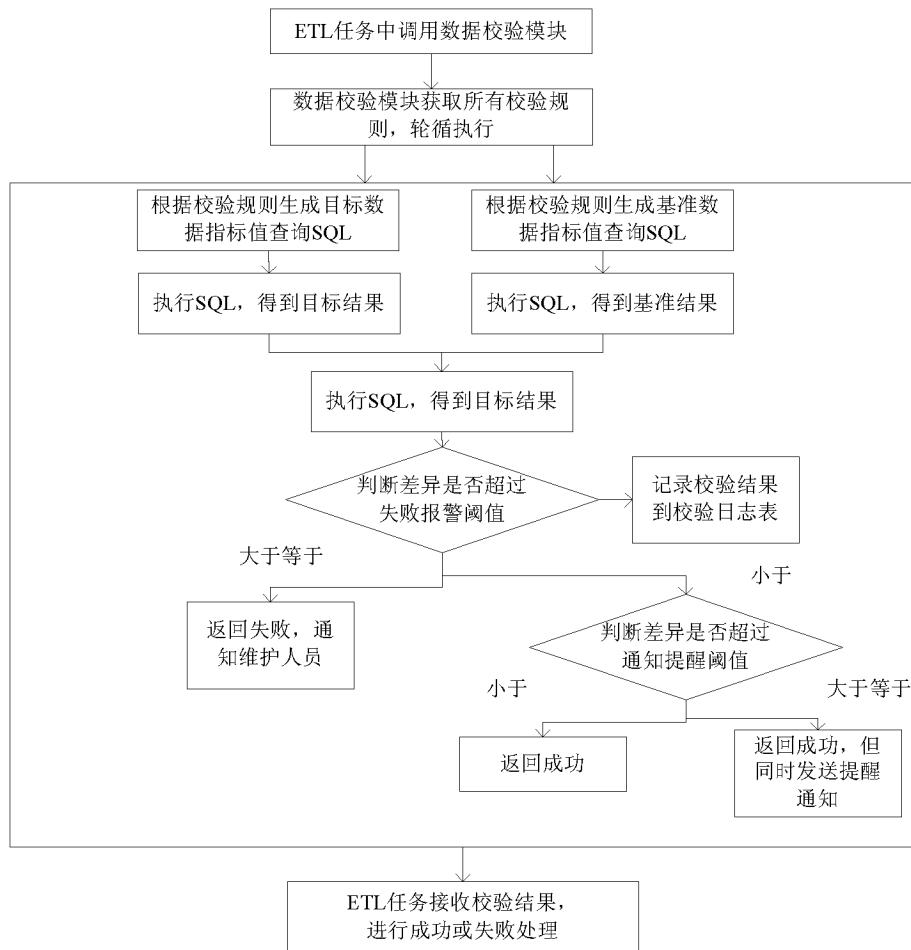


图 9