

一种基于集团管理多维数据高效处理方法

技术领域

本发明涉及医疗管理领域，尤其涉及一种基于集团管理多维数据高效处理方法。

背景技术

现有的集团化信息管理系统中，通常涉及各业务端口的对接、多源数据的采集与数据共享的操作，其中各业务端口所涉及的多源数据将分布在不同业务系统以及不同类型的数据库中，在数据的调用与实际使用过程中，由于数据的源头不同，将导致数据结构，数据格式以及数据的标准化程度不同，无法对采集的数据更好地进行结构化与标准化处理，现有的数据处理技术中，通常采用数据采集即处理的方案，待数据处理完成后再进行统一的数据抽取与调用，但是这种处理方式对于多数据库批量数据的处理效率较慢，严重影响数据更新量较大的信息管理系统的运行效率。

公开号为 CN102650995A 的中国专利，提供了一种多维数据分析模型生成系统及方法，此专利虽然也涉及多维数据的处理方法的设计，但是此专利是通过建立可复用的多维数据分析模型进行数据处理的，其对多维数据的处理主要用以提高多维数据处理的精确性，但是此专利通过建立多个多维数据模型分别进行数据数据后，再进行各个模型之间的关联与组合，无法进一步提高多维数据的处理效率，从而无法进一步解决数据更新量较大的信息管理系统的运行效率的优化问题。公开号为 CN103488475B 的中国专利，提供了一种多维数据分析系统和多维数据分析方法，此专利涉及多维数据的分析方法用以提高用户操作的易用性，但是此专利是通过将接收的数据直接划为不同维度下的路径信息，根据路径信息进行数

据的分别传输与分析。

因此，针对现有多维数据高效处理过程中存在的问题，本发明中提供了一种基于集团管理多维数据高效处理方法。

发明内容

针对上述存在的问题，本发明中提供了一种基于集团管理多维数据高效处理方法，具体包括，建立了多数据库下多维管理数据的并行采集方法，对采集的多维管理数据进行多源数据库表的数据整合，将整合后的数据库表进行数据标准化处理，在业务系统数据库中调用标准化处理后的数据，并根据业务系统中的业务规则进行数据调用与查询。

优选的，所述的数据调用与查询，包括数据预警，数据通知以及标准化量化数据显示。

优选的，所述的多数据库包括 MySQL，Oracle 和 Excel 数据库，所述的并行采集方法中，通过建立多源数据库表缓存区，在多源数据库表缓存区中对采集的多维管理数据进行数据整合与数据标准化处理。

优选的，所述的并行采集中，采用基于 python 的 pandas 工具建立与多数据库的连接属性；所述的连接属性包括数据库表地址，自定义登录名称与登陆密码，标志位判定。

优选的，所述的数据标准化处理的具体步骤为，采集多数据库下多维管理数据的元数据，此时并未对元数据进行数据，待基于多源数据库表的数据整合完成后，再进行数据的统一标准化处理。

优选的，所述的多源数据库表的数据整合的具体步骤为：

A1、对多数据库下的数据库表进行标签化分类，根据所述的标签化分类中的

标签进行数据库表的聚类处理；

A2、在 A1 的基础上，将所述的聚类处理分为并列标签聚类与树状结构标签聚类；

A3、在进行多数据库下多维管理数据的并行采集时，通过循环遍历的方式分别根据标签对并列标签聚类与树状结构聚类进行分类处理；

A4、在 A3 的基础上，进行多数据库下数据库表的合并，在合并后的数据库表上进行数据的标准化处理，所述的数据标准化处理包括数据清洗，数据去重，数据格式统一化，异常数据值剔除。

优选的，所述的并列标签聚类与树状结构标签聚类，根据业务规则对数据库表中的标签进行分级，在多数据库下的数据库表中将同一级别的标签划分于同一集合中。

优选的，在所述的树状结构标签聚类中，将各级别标签的属性关系划分于一个集合中。

优选的，所述的数据标准化处理，建立基于业务系统中所需数据的统一调用逻辑，所述的统一调用逻辑用以数据在业务系统中应用。

优选的，所述的的数据标准化处理，将处理后的数据存储于多源数据库表缓存区中，通过 PostgreSQL 数据库对多源数据库表缓存区中的数据进行调用。

与现有技术相比，本发明的有益效果在于：

(1) 本发明中所述的一种基于集团管理多维数据高效处理方法，通过建立基于 python 的跨数据库的数据抽取与数据标准化处理方法，用以进一步提高多源数据的采集与集团管理系统中，业务系统下多业务端口下数据共享的效率。

(2) 在 (1) 的基础上，本发明中所述的业务系统采用的是 MySQL 数据库，

用以避免其他数据库在进行数据迁入和系统历史 bug 等原因导致噪声数据的引入，从而对集团管理多维数据高效处理系统带来运行干扰，需要额外的增设专门的数据清洗流程，影响数据处理的效率。

(3) 在(2)的基础上，本发明需要将这三类数据源的数据通过 python 脚本统一抽取并存储于 PostgreSQL 数据库中，根据业务规则对标准化后的数据统一进行数据清洗操作，从而进行数据的筛选，数据修复操作，保证数据在业务系统中展示的精确性。

(4) 在(3)的基础上，在完成统一的数据清洗工作后，本发明根据集团管理业务的实际业务需求，利用 pandas 工具进行深度加工和处理，最终形成符合业务需求，便于查询的数据结构，存储在各类新的数据表中，最后，由业务系统调用该库的数据，实现低成本的数据分析和查询系统，本系统可以承载集团三百多张数据库表，最大业务表百亿行真实业务数据的处理和查询，效率高，成本低。

附图说明

图 1 为一种基于集团管理多维数据高效处理方法流程图。

具体实施方式

实施例 1:

本实施例中所述的一种基于集团管理多维数据高效处理方法，如图 1，建立了多数据库下多维管理数据的并行采集方法，对采集的多维管理数据进行多源数据库表的数据整合，将整合后的数据库表进行数据标准化处理，在业务系统数据库中调用标准化处理后的数据，并根据业务系统中的业务规则进行数据调用与查询。

所述的并行采集中，采用基于 python 的 pandas 工具建立与多数据库的连接

属性；所述的连接属性包括数据库表地址，自定义登录名称与登陆密码，标志位判定。

具体的，所述的数据抽取的具体流程为：

S1、建立多源数据库表缓存区，定义基于 python 的多数据源下的通信接口，配置基于 python 与 Oracle 的连接属性，并选择 oracle 的配置端口，在此基础上，获取 oracle 中数据库表的名称与游标信息，以此调用 oracle 中所需的数据信息后关闭游标，将所调用的数据信息存储于多源数据库表缓存区中；

S2、同时配置基于 mysql 的连接属性，待与 mysql 连接成功后利用 python 中的 pandas 工具进行 mysql 中指定数据的定位与批量导入操作，同时指定待导入数据的数据格式；

S3、在 S2 的基础上，所述的数据格式包括 Excel 格式数据，其中所述的 Excel 格式数据的导入方式为，指定 Excel 表单类型，并获取指定表单类型中的行数据与列数据的数量，通过行数据与列数据建立索引列表，通过索引列表逐个读取 Excel 表单中对应的数据。

所述的数据标准化处理，建立基于业务系统中所需数据的统一调用逻辑，所述的统一调用逻辑用以数据在业务系统中应用。

所述的业务规则约束中，根据业务规则标准化后的数据进行统一的数据清洗，在所述的数据清洗过程中，如订单总金额与支付方式明细汇总不一致的，根据收费项目明细进行佐证，尝试修复数据。如果无法修复的，这类数据会进行丢弃，并提交清单给到业务部门进行核实和修正，如果数据得到修正，会在业务系统中完成更新，并通过抽取逻辑更新到统一的数据库中。

权利要求书

1、一种基于集团管理多维数据高效处理方法，其特征在于，具体包括，建立了多数据库下多维管理数据的并行采集方法，对采集的多维管理数据进行多源数据库表的数据整合，将整合后的数据库表进行数据标准化处理，在业务系统数据库中调用标准化处理后的数据，并根据业务系统中的业务规则进行数据调用与查询。

2、根据权利要求1所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的数据调用与查询，包括数据预警，数据通知以及标准化量化数据显示。

3、根据权利要求1所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的多数据库包括MySQL，Oracle和Excel数据库，所述的并行采集方法中，通过建立多源数据库表缓存区，在多源数据库表缓存区中对采集的多维管理数据进行数据整合与数据标准化处理。

4、根据权利要求3所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的并行采集中，采用基于python的pandas工具建立与多数据库的连接属性；所述的连接属性包括数据库表地址，自定义登录名称与登陆密码，标志位判定。

5、根据权利要求1所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的数据标准化处理的具体步骤为，采集多数据库下多维管理数据的元数据，此时并未对元数据进行数据，待基于多源数据库表的数据整合完成后，再进行数据的统一标准化处理。

6、根据权利要求5所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的多源数据库表的数据整合的具体步骤为：

A1、对多数据库下的数据库表进行标签化分类，根据所述的标签化分类中的标签进行数据库表的聚类处理；

A2、在 A1 的基础上，将所述的聚类处理分为并列标签聚类与树状结构标签聚类；

A3、在进行多数据库下多维管理数据的并行采集时，通过循环遍历的方式分别根据标签对并列标签聚类与树状结构聚类进行分类处理；

A4、在 A3 的基础上，进行多数据库下数据库表的合并，在合并后的数据库表上进行数据的标准化处理，所述的数据标准化处理包括数据清洗，数据去重，数据格式统一化，异常数据值剔除。

7、根据权利要求 6 所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的并列标签聚类与树状结构标签聚类，根据业务规则对数据库表中的标签进行分级，在多数据库下的数据库表中将同一级别的标签划分于同一集合中。

8、根据权利要求 7 所述的一种基于集团管理多维数据高效处理方法，其特征在于，在所述的树状结构标签聚类中，将各级别标签的属性关系划分于一个集合中。

9、根据权利要求 1 所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的数据标准化处理，建立基于业务系统中所需数据的统一调用逻辑，所述的统一调用逻辑用以数据在业务系统中应用。

10、根据权利要求 5 所述的一种基于集团管理多维数据高效处理方法，其特征在于，所述的的数据标准化处理，将处理后的数据存储于多源数据库表缓存区中，通过 PostgreSQL 数据库对多源数据库表缓存区中的数据进行调用。