



(12) 发明专利

(10) 授权公告号 CN 101197876 B

(45) 授权公告日 2012. 02. 29

(21) 申请号 200610157411. 7

(22) 申请日 2006. 12. 06

(73) 专利权人 中兴通讯股份有限公司

地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦A座6层

(72) 发明人 陈俊 蒋柯 丁岩 杨淑艳

(74) 专利代理机构 深圳鼎合诚知识产权代理有限公司 44281

代理人 薛祥辉

(51) Int. Cl.

H04L 12/24 (2006. 01)

G06F 17/30 (2006. 01)

(56) 对比文件

CN 1617138 A, 2005. 05. 18, 全文.

CN 1556611 A, 2004. 12. 22, 摘要、说明书第 2-3 页.

孙泳、刘少辉、史忠植. 数据仓库中多维分析的数据展现. 计算机工程与应用. 2004, 174-177.

李林花、钱越英. 数据仓库多维分析模型的设计. 计算机工程与应用. 2004, 185-187.

审查员 吴翔晖

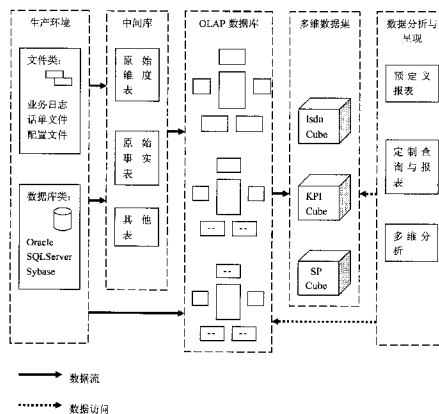
权利要求书 2 页 说明书 8 页 附图 3 页

(54) 发明名称

一种对消息类业务数据进行多维分析的方法和系统

(57) 摘要

本发明是一种对消息类业务数据进行多维分析的方法和系统,其步骤包括、数据采集步骤、数据整合步骤,数据转换与装载模块负责将中间数据库的数据转换处理到一个具有星型架构特征的关系型数据库中,对消息类业务的各种待分析数据进行整理,将具有共同或相近特性的数据进行归类,并纳入到一个统一的维度层次结构中;对中间数据库的数据进行进一步的清洗、转换以处理到另一个多维数据库中;及数据分析与呈现步骤。对应的该系统,包括独立设置数据采集模块,数据转换与装载模块及数据分析与呈现模块,采用本发明,能在相似的增值业务系统中共享多维数据模型,减少系统的开发周期和成本。应用该模型建立的系统所具有的多维度数据分析的功能,可为运营商和设备制造商提供更为灵活方便的数据统计方法及决策支持。



CN 101197876 B

1. 一种对消息类业务数据进行多维分析的方法,其特征在于,该方法包括如下步骤;

1) 数据采集,数据采集模块负责将业务系统中分散在各个网元节点上的生产数据采集到统计分析服务器上的一个中间数据库中;

2) 数据整合,数据转换与装载模块负责将中间数据库的数据转换处理到一个具有星型架构特征的关系型数据库中,并根据需要进一步处理到数据库中,整合包括:对消息类业务的各种待分析数据进行整理,将具有共同或相近特性的数据进行归类,并纳入到一个统一的维度层次结构中;对中间数据库的数据进行进一步的清洗、转换以处理到另一个多维数据库中;所述数据整合步骤中,设计星型架构作为数据组织的基本方式,即以事实数据表为中心,多个维度表通过各自的键列联接到事实数据表上;其中,根据星型架构设计与之对应的数据仓库架构,将该多维数据库采用增量更新和按分区刷新相结合的方式装载到数据仓库的 Cube(多维立方体)中,所述星型架构所包含的事实表包含条数、字节流量和费用字段,且每个星型架构所包含的维度表都有且仅有一个键列联接到事实数据表,中间库中维度数据转换处理包括:

维度表数据同步更新:在 DTS(数据转换服务)包里完成配置信息的初始化;建立到中间库的连接;读取中间库号码表的标志列确定本次待处理行;通过数据库基本操作指令实现维度数据的同步;将原始维度表异常数据写入异常维度信息表;更新关系型数据库维度表的属性字段信息;

中间库中事实表数据抽取与转换:

将中间库中分散在多表的原始事实数据进行规整处理,汇集到关系型数据库的一个临时表中,该临时表具备对应的星型架构事实表和维度表在最细粒度下的聚合信息;对该临时表所包含的所有维度信息进行处理,以使其与关系型数据库的维度表在最低的层次上发生关联,并经过中间转换;增量更新到关系型数据库的事实表中,通过处理将原始事实表中所有具体的维度取值转换为维度表中对应的键值;规整处理的方法主要包括库表之间的关联更新和聚合操作;中间库中零散的事实数据在多次转换后填充到星型架构的事实表中。

3) 数据分析与呈现,通过数据分析与呈现模块负责对数据仓库或关系型数据库中具有多维特征的业务数据进行分析 and 呈现。

2. 如权利要求 1 所述的对消息类业务数据进行多维分析的方法,其特征在于在数据采集步骤中,首先定义针对不同数据源的数据抽取标准方法,完成异构数据源的原始数据采集工作,并保存到本地的一个中间数据库中,该中间数据库包含了业务系统所有经初步清洗的待分析数据;在原始数据采集过程中通过配置文件的方式定义待抽取库表的详细信息,包括表名、字段名、字段类型。

3. 如权利要求 1 所述的对消息类业务数据进行多维分析的方法,其特征在于,所述维度数据的同步包括;

A 新增维度数据处理:根据源表的最新 id 值将本次新增维度数据插入到关系型数据库维度表;

B 维度信息更新处理:根据源表的 id 值确定本行是不是新增维度信息,根据源表的标志列确定本行的部分信息发生了更新,在对关系型数据库中的维度表进行更新操作中关联已确定发生了信息更新的记录以更新维度表数据;

C 维度表属性字段信息通过一组存储过程来实现,以完成维度成员属性的计算。

4. 一种对消息类业务数据进行多维分析的系统,其特征在于该系统包括独立设置的以下模块:

数据采集模块,负责将业务系统中分散在各个网元节点上的生产数据采集到统计分析服务器上的一个中间数据库中,具体包括文件类生产数据处理子模块和数据库类生产数据处理子模块,文件类型的生产数据采用两个独立进程分别完成文件获取与入库工作;数据库类生产数据采用链接服务器、DTS 和数据库复制三种方式同步数据;

数据转换与装载模块,负责将中间数据库的数据转换处理到一个具有星型架构特征的关系型数据库中,并根据需要进一步处理到数据仓库中,数据转换具体包括三个主要模块;维度表处理模块、事实表处理模块和异常数据处理模块;维度表处理模块负责完成维度数据的收集和更新,以及对维度的属性数据进行预先计算处理;事实表处理模块负责完成事实数据的收集和更新,以及通过维护与维度表的主外键关系实现与维度表的关联;异常数据处理模块负责对系统异常数据进行额外处理;数据装载模块负责将该星型架构的多维数据库即事实表、维度表通过增量更新和按分区刷新相结合的方式装载到 Cube 中;

数据分析与呈现模块,负责对数据仓库或关系型数据库中具有多维特征的业务数据进行分析 and 呈现,具体包括预定义统计报表子模块、定制查询与报表子模块以及多维分析子模块;预定义报表子模块是将查询主体中行列定义、字段具体含义均预先定义为报表定义文件,定制查询子模块是将与报表的行列字段定义在界面中预定义,报表子模块储存各种报表文件,多维分析子模块,其是展现所有 Cube 维度及度量值的要素,用户首先选定某些维度的取值,之后在界面上设置一个或多个维度作为分析列、一个或多个维度作为分析行、一个或多个度量值作为数据项,通过动态展开分析行及分析列可以获得在不同维度取值影响下的度量值数据,分析行及分析列维度在不同层次(level)上自动进行数据的聚合;定制报表的数据源既包括关系型数据库,也包括数据仓库,而多维分析模块的数据源只基于数据仓库的 Cube。

## 一种对消息类业务数据进行多维分析的方法和系统

### 技术领域

[0001] 本发明属于电信技术领域,涉及一种对消息类增值业务数据进行分析的方法和系统。

### 背景技术

[0002] 增值业务是电信业一个重要的利润支撑点,消息类业务是电信增值业务中的重要收入组成部分。对增值业务的各类数据进行有效分析,将有助于指导业务开展、提高服务质量、提供决策支持、最终实现持续的利润增长,因此数据分析的重要性毋庸置疑。

[0003] 现阶段增值业务的数据分析手段,通常包括两大类:

[0004] 一类是增值业务设备厂商提供的 O&M(运维系统)的分析手段,通过本地操作维护系统可以实现特定业务流程的信令跟踪、业务观察、性能计数器、统计报表等功能,这些功能主要用于进行问题定位、原始数据收集与汇总;

[0005] 一类是增值业务的 KPI 数据采集及上报,比如通过 SNMP(简单网管协议)、FTP、FTAM(文件传输访问和管理)等接口提供必要的规范化设备侧信息,交由运营商或者第三方进行后续汇总分析处理,这一类总的说来属于增值业务的北向数据接口。对于这种情况,通常是运营商制定统计指标规范,并由设备提供商按规范进行实现。

[0006] 现实的情况是,O&M 系统虽然可以深入到业务流程的内部,获得最详细的业务数据,但 O&M 更多关注的是特定的业务流程性信息,而北向数据接口由于受到运营商规范的限制,往往提供的是硬性裁剪后的二维数据,偏重于宏观的 KPI 指标数值,不足以揭示业务的各种丰富的内在联系。即使运营商通过 BOSS 系统或其它第三方系统对采集的原始 CDRs 进行后续处理,例如经营分析系统,也通常会缺失业务系统中有价值的原始信息。

[0007] 总的来讲,现有的消息类增值业务的数据统计分析系统和方法具有以下弊端:

[0008] 1. 相似的增值业务系统缺少统一的数据分析模型,系统重复建设,增大了开发工作量。

[0009] 2. 基于二维的 RDBMS(关系型数据库),业务数据的内在复杂的关联无法很好的体现,不利于实现联机事务分析和决策支持。

### 发明内容

[0010] 为解决所述技术问题,本发明提供一种对消息类业务数据进行多维分析的方法和系统,该方法和系统能够建立简洁、快速的数据分析模型,减少开发工作量,并有利于联机支持。

[0011] 本发明的另一个目的在于提供一种对消息类业务数据进行多维分析的方法和系统,该方法和系统通过星形架构的数据组织模式建立多维数据模型,从而能够快速地进行数据分析。

[0012] 本发明所涉及的一种对消息类业务数据进行多维分析的方法包括如下步骤:

[0013] 1、数据采集,数据采集模块负责将业务系统中分散在各个网元节点上的生产数据

采集到统计分析服务器上的一个中间数据库中；

[0014] 2、数据整合,数据转换与装载模块负责将中间数据库的数据转换处理到一个具有星型架构特征的关系型数据库中,并根据需要进一步处理到数据库中,整合包括:对消息类业务的各种待分析数据进行整理,将具有共同或相近特性的数据进行归类,并纳入到一个统一的维度层次结构中;对中间数据库的数据进行进一步的清洗、转换以处理到另一个多维数据库中;所述数据整合步骤中,设计星型架构作为数据组织的基本方式,即以事实数据表为中心,多个维度表通过各自的键列联接到事实数据表上;其中,根据星型架构设计与之对应的数据仓库架构,将该多维数据库采用增量更新和按分区刷新相结合的方式装载到数据仓库的 Cube(多维立方体)中,所述星型架构所包含的事实表包含条数、字节流量和费用字段,且每个星型架构所包含的维度表都有且仅有一个键列联接到事实数据表,中间库中维度数据转换处理包括:

[0015] 维度表数据同步更新:在 DTS(数据转换服务)包里完成配置信息的初始化;建立到中间库的连接;读取中间库号码表的标志列确定本次待处理行;通过数据库基本操作指令实现维度数据的同步;将原始维度表异常数据写入异常维度信息表;更新关系型数据库维度表的属性字段信息;

[0016] 中间库中事实表数据抽取与转换:

[0017] 将中间库中分散在多表的原始事实数据进行规整处理,汇集到关系型数据库的一个临时表中,该临时表具备对应的星型架构事实表和维度表在最细粒度下的聚合信息;对该临时表所包含的所有维度信息进行处理,以使其与关系型数据库的维度表在最低的层次上发生关联,并经过中间转换;增量更新到关系型数据库的事实表中,通过处理将原始事实表中所有具体的维度取值转换为维度表中对应的键值;规整处理的方法主要包括库表之间的关联更新和聚合操作;中间库中零散的事实数据在多次转换后填充到星型架构的事实表中;

[0018] 3、数据分析与呈现,通过数据分析与呈现模块负责对数据仓库或关系型数据库中具有多维特征的业务数据进行分析 and 呈现。

[0019] 在数据采集步骤中,首先定义针对不同数据源的数据抽取标准方法,完成异构数据源(各类数据库、话单文件和业务日志文件)的原始数据采集工作,并保存到本地的一个中间数据库中,该中间数据库包含了业务系统所有经初步清洗的待分析数据。在原始数据采集过程中通过配置文件的方式定义待抽取库表的详细信息,包括表名、字段名、字段类型等。

[0020] 所述维度数据的同步包括:

[0021] A 新增维度数据处理:根据源表的最新 id 值将本次新增维度数据插入到关系型数据库维度表;

[0022] B 维度信息更新处理:根据源表的 id 值确定本行是不是新增维度信息,根据源表的标志列确定本行的部分信息发生了更新,在对关系型数据库中的维度表进行更新操作中关联已确定发生了信息更新的记录以更新维度表数据;

[0023] C 维度表属性字段信息通过一组存储过程来实现,以完成维度成员属性的计算。

[0024] 根据该方法建立的一种对消息类业务数据进行多维分析系统,包括以下模块:

[0025] 数据采集模块,负责将业务系统中分散在各个网元节点上的生产数据采集到统计

分析服务器上的一个中间数据库中,具体包括文件类生产数据处理子模块和数据库类生产数据处理子模块,文件类型的生产数据采用两个独立进程分别完成文件获取与入库工作;数据库类生产数据采用链接服务器、DTS 和数据库复制三种方式同步数据;

[0026] 数据转换与装载模块,负责将中间数据库的数据转换处理到一个具有星型架构特征的关系型数据库中,并根据需要进一步处理到数据仓库中,数据转换具体包括三个主要模块:维度表处理模块、事实表处理模块和异常数据处理模块;维度表处理模块负责完成维度数据的收集和更新,以及对维度的属性数据进行预先计算处理;事实表处理模块负责完成事实数据的收集和更新,以及通过维护与维度表的主外键关系实现与维度表的关联;异常数据处理模块负责对系统异常数据进行额外处理;数据装载模块负责将该星型架构的多维数据库即事实表、维度表通过增量更新和按分区刷新相结合的方式装载到 Cube 中;

[0027] 数据分析与呈现模块,负责对数据仓库或关系型数据库中具有多维特征的业务数据进行分析 and 呈现,具体包括预定义统计报表子模块、定制查询与报表子模块以及多维分析子模块;预定义报表子模块是将查询主体中行列定义、字段具体含义均预先定义为报表定义文件,定制查询子模块是将与报表的行列字段定义在界面中预定义,报表子模块储存各种报表文件,多维分析子模块,其是展现所有 Cube 维度及度量值的要素,用户首先选定某些维度的取值,之后在界面上设置一个或多个维度作为分析列、一个或多个维度作为分析行、一个或多个度量值作为数据项,通过动态展开分析行及分析列可以获得在不同维度取值影响下的度量值数据,分析行及分析列维度在不同层次(level)上自动进行数据的聚合;定制报表的数据源既包括关系型数据库,也包括数据仓库,而多维分析模块的数据源只基于数据仓库的 Cube。

[0028] 采用本发明,能在相似的增值业务系统中共享多维数据模型,减少系统的开发周期和成本。应用该模型建立的系统所具有的多维度数据分析的功能,可为运营商和设备制造商提供更为灵活方便的数据统计方法及决策支持。

#### 附图说明

[0029] 图 1 是本发明业务数据分析系统的模块构成与其它元件的连接示意图;

[0030] 图 2 是本发明业务数据分析系统的业务数据星型架构示意图;

[0031] 图 3 是本发明业务数据分析系统的数据转换与装载模块工作原理示意图。

#### 具体实施方式

[0032] 如图 1 所示,本发明的业务统计分析系统主要有以下三个模块组成,包括数据采集模块、数据转换与装载模块、数据分析与呈现模块。

[0033] 数据采集模块负责将业务系统各业务处理机的话单文件、日志文件和配置文件等获取到统计服务器上并入库,同时也将业务系统各节点数据库的用户信息、原始配置、计数器数据获取到统计服务器的数据库中。

[0034] 数据转换与装载模块负责将已经入库的各类原始数据进行进一步清洗和转换,使之符合星型架构的库表结构,并需要根据按照预先设计的规则将星型架构的关系型数据装载入数据仓库。

[0035] 设计星型架构(图 2 所示)作为数据组织的基本方式,即以事实数据表为中心,多

个维度表通过各自的键列联接到事实数据表上。根据业务数据自身特点及方便后续分析的考虑,为消息类业务设计多个这样的星型架构。

[0036] 各星型架构的事实表的数据不重复,以保证整个业务系统数据的一致性,维度表则为所有事实表所公用。该多维数据模型覆盖业务系统中所有具有统计分析价值的数据库。

[0037] 通过对消息类业务的各种待分析数据库进行整理,将具有共同或相近特性的数据库进行归类,并纳入到一个统一的维度层次结构中。消息类业务至少可以归纳出以下维度:用户终端号码、用户终端型号、时间、服务提供商信息、业务应用类型、业务计费类型、OT 话单类型、地区信息、业务状态码信息等。

[0038] 对于每一个维度,进一步设计其层次结构级别和维度的成员属性,例如对于用户终端号码维度,设计为包括归属省份代码、地区代码、服务品牌代码等三个级别,在每一个级别上根据需要再设计出省份描述、地区描述、品牌编码和品牌描述等维度成员的属性。

[0039] 星型架构所包含的事实表包含条数、字节流量和费用字段,且每个星型架构所包含的维度表都有且仅有一个键列联接到事实数据库。条数、字节流量和费用等作为业务多维分析的最重要的度量值,后续的统计分析都在计算它们的具体数值、研究数值的变化规律、影响因素、未来走向等。

[0040] 对中间数据库的数据库进行进一步的清洗、转换以处理到另一个多维数据库中,该多维数据库符合前述的星型架构模型。

[0041] 根据星型架构设计与之对应的数据仓库架构,将该多维数据库采用增量更新和按分区刷新相结合的方式装载到数据仓库的 Cube(多维立方体)中。

[0042] 数据分析与呈现模块负责对装载入数据仓库的数据库和某些原始关系型数据库进行分析和呈现,该模块为用户提供统计报表、即时查询和多维分析的人机界面。

[0043] 并根据每个 Cube 所装载的业务数据库的特点,定制不同主题的统计报表。

[0044] 再根据联机分析主题设计 B/S 界面,完成界面分析主题翻译成数据仓库操作指令,进而将查询返回结果以报表或图表的形式进行展现。

[0045] 以上三个模块可以放置于一台计算机上,也可以根据系统业务量负荷分设在不同的计算机上。

[0046] 以下分别对业务统计分析系统各个模块的工作原理进行说明:

[0047] 1、数据采集模块

[0048] 数据采集模块负责将业务系统中分散在各个节点上的生产数据库收集到上。分析服务器这里所指的节点包括多个具有同等业务处理功能的业务处理机,以及话单服务器、日志服务器和操作维护服务器等。无论是 MMS 彩信业务还是 SMS 短消息业务,从功能划分上都具有这些服务器,其操作系统包括 Unix、Linux 和 Windows 等主流操作系统。各节点的数据存放的形式包括数据库(Oracle/Sybase/SQL Server)、文件(话单文件、业务日志文件、配置文件等)。

[0049] 数据采集模块由独立的一组进程和基于数据库的一组任务来完成数据库的采集。数据库收集的手段包括文件获取及入库、基于链接服务器和数据库复制技术的数据库间数据库传送。对于文件类的生产数据库由该组独立进程负责采集,对于数据库类的生产数据库由该组基于数据库的任务负责采集。该模块包括文件类生产数据库处理子模块和数据库类生产数据库处理子模块,文件类生产数据库采用两个独立进程分别完成文件获取与入库工作,进程间不进

行通讯。

[0050] 以下说明文件类生产数据的采集处理过程：

[0051] 统计服务器上包括两个独立的与文件生产数据采集相关的进程，一个是文件获取进程，一个是文件解析入库进程。

[0052] 文件获取进程首先读取包括各节点 IP 地址、口令、文件存放路径等配置信息，之后启动 FTP 的客户端，将各业务节点的文件收集到统计服务器的指定目录下。文件获取进程所依赖的服务器地址信息、日志及话单文件的存放路径信息、文件的格式（文本或者二进制）、字段分隔符、文件的命名规则及扩展名信息、统计分析服务器上日志文件的存放路径等众多的信息都是可配置的。本进程的不同线程完成不同类型文件的处理。对于同类文件的处理又主要分为两个阶段，首先是文件获取到本地一个用于中转的文件夹“中间数据库 (Stagedb)”（以下均称为“中间库”）中，然后根据配置文件对本类文件进行初步分拣和信息完整性检查，对于不符合规范的错误日志将筛选并保存到独立的文件中以备后续进行额外处理。经过以上处理的文件按照文件的不同种类被放置于指定文件夹“关系型数据库 (olapdb)”。

[0053] 文件解析入库进程负责将已获取到分析服务器上的文件写入数据库。入库过程对于不同的目标数据库类型具有相似的处理方式。一个处理循环包括：读取各类配置信息、将待入库数据读入内存、对源文件进行改名、创建对指定目标数据库的数据库连接、组装 SQL 语句、完成批量入库、对已处理文件改名、重新载入新文件。

[0054] 对于存放在各业务节点数据库中的数据，由于生产数据库包括 Oracle、Sybase 和 SQL Server 等，因此数据抽取方法不一样，本系统定义了两种数据抽取标准，对于 Oracle 和 Sybase，采用链接服务器的方式；对于 SQL Server，则采用数据库复制和 DTS（数据转换服务）的方式。

[0055] 无论是文件方式还是数据库方式，生产数据最终都被综合到 stagedb 中，该中间库包含了所有的经初步抽取和格式转换后的原始业务信息，也包括所有有用的系统配置信息。

[0056] 对于系统注册用户数据、手机号段配置信息、SP 业务信息、终端型号配置等信息，数据分布在业务系统的各个节点数据库中，其数据导入的基本过程以系统注册用户数据为例，包括：

[0057] 1). 制定对生产数据库的访问通道。读取配置文件和数据配置表以确定用户数据所在的服务器局号、节点号、IP 地址、数据库名称、访问方式等，保存为中间变量。如果生产库位于多个节点，则循环读入上述信息。根据中间变量动态创建链接服务器以链接至对应的数据库，并测试链接的可用性；此处链接服务器根据源数据库的类型不同，设定不同的创建参数，包括使用不同数据库提供程序如 Oracle Provider for OLE DB 或 Microsoft OLE DB Provider for ODBC 等。

[0058] 2). 完成对生产数据库的数据采集和初步整理。建立好对各生产库的数据库访问通道后，对待抽取生产库表进行顺序读取。不同特性的生产库表在数据量上差别很大，根据待抽取数据库表的 id 主键取值确定本条记录是否为新增记录，根据最后更新时间字段确定本条记录是否进行过更新，并通过 DTS（数据转换服务）或者存储过程完成新增及修改记录向中间库的同步。



## [0059] 2、数据转换与装载模块

[0060] 图 3 是数据转换与装载模块的示意图。本模块的基本功能是对已经统一写入中间数据库 stagedb 的数据进行进一步清洗转换处理,处理的结果数据存放在 olapdb 的数据库中,该数据库的多表间具有星型架构(如图 2 所示)的特点,为装载入数据仓库做好准备。之后,按照已经设计完成的数据仓库架构,定时任务会触发从星型架构的 olapdb 表往数据仓库装载多维数据的过程。

[0061] 从 stagedb 到 olapdb 的数据转换处理主要包含以下过程:

[0062] 1) 维度表数据同步更新;

[0063] stagedb 库包含了业务系统各种维度的原始数据,但这些数据是零散的,且通常不符合维度设计的层次结构要求,因此通过 DTS(数据转换服务)和存储过程来实现生产表到维度表的转换。考虑到每次维度数据同步更新涉及到的数据量可能较大,加上维度表与事实表数据之间有主外键关联关系,维度表数据不采用每次重建的方式,而是增量更新的方式。以用户号码维度的转换过程为例,处理流程依次为:在 DTS 包里完成配置信息的初始化;建立到中间库的连接;读取中间库号码表的标志列确定本次待处理行;通过数据库基本操作指令实现维度数据的同步;将原始维度表异常数据写入异常维度信息表;更新 olapdb 维度表的属性字段信息。

[0064] 处理流程中所述的维度数据的同步包括:

[0065] A 新增维度数据处理:根据源表的最新 id 值将本次新增维度数据插入到 olapdb 维度表;

[0066] B 维度信息更新处理:根据源表的 id 值确定本行不是新增维度信息,根据源表的标志列确定本行的部分信息发生了更新,在对 olapdb 数据库中的维度表进行更新操作中关联已确定发生了信息更新的记录以更新维度表数据;

[0067] 维度表属性字段信息通过一组存储过程来实现,以完成诸如注册用户数、新增注册用户数等维度成员属性的计算。

[0068] 经过以上处理,将 stagedb 库的新增及变化了的原始维度数据以较小的计算量更新到 olapdb 的维度表中。

[0069] 2) 事实表数据抽取与转换;

[0070] 一个事实表数据抽取与转换的过程包括:

[0071] 首先将 stagedb 中分散在多表的原始事实数据进行规整处理,汇集到 olapdb 的一个临时表中,该临时表具备对应的星型架构事实表和维度表在最细粒度下的聚合信息,之后对该临时表所包含的所有维度信息进行处理,以使其与 olapdb 的维度表在最低的层次(Level)上发生关联,并经过若干中间转换,最后增量更新到 olapdb 的事实表中。例如,一个话单原始事实表中包含了一个日期类型的时间字段,而在话单事实表中表示该时间的字段是一个 Int 型字段,该字段与时间维度表的 Id 字段具有外键关联关系。从原始事实表转换为事实表后,原来的 datetime 类型的时间信息就变为对应时间维度表的一个外键值。通过一系列处理将原始事实表中所有具体的维度取值转换为维度表中对应的键值。规整处理的方法主要包括库表之间的关联更新和聚合操作(group by)。Stagedb 中零散的事实数据在多次转换后填充到星型架构的事实表中。

[0072] 由于业务数据的数据量设计为忙时 500 条/秒或更多,因此在按星型架构组织的

数据库 olapdb 中,事实表数据采用按月分表存储。

[0073] 3) 质疑数据处理;

[0074] 对于原始事实表中某些无法与对应的维度表产生主外键关联关系的记录,称为质疑数据。这类数据将一直被保存在异常数据表中,并在每次的事实表数据抽取与转换处理的最后阶段进行额外处理,重新试图与维度表进行匹配。若匹配上,则重新处理到事实表中,否则一直被保留在异常数据表中。

[0075] 数据装载主要包括以下过程:

[0076] 基于星型架构的事实表和维度表构建 Cube(数据仓库立方体),根据后续 OLAP(联机分析与处理)分析的热点设计聚合度,增量处理维度更新和维度属性值,根据 processed 标识字段增量装载事实表数据至 Cube 的对应分区。

[0077] 数据装载具有以下特点:

[0078] A、Cube 所依赖的事实表不是基于表,而是基于由多个分表所组成的联合视图。

[0079] B、Cube 采用 MOLAP(多维 OLAP)的存储方式,同时由于大量的数据分析通常发生在少数几个连续的自然月时间粒度里,因此 Cube 按月进行自动分区,以加快后续 OLAP 分析时的相应速度。

[0080] 3、数据分析与呈现模块

[0081] 数据分析与呈现模块直接面向最终用户,为用户提供直观、方便的 B/S 界面来分析业务系统的数据。该模块所基于的数据以数据转换与装载模块中生成的数据仓库为主,也可以包括关系型数据库的数据。

[0082] 数据分析与呈现模块包括预定义统计报表子模块、定制查询报表子模块以及多维分析子模块。其中,报表子模块根据数据仓库 Cube 的划分,又分为多个报表大类,包括综合类、终端类和 SP(服务提供商)类等。预定义报表可以支持众多数据的统计分析,可采用小时报表、日报表和月报表的形式进行展现,提供几大类、几十个小类和数百个统计项,典型的统计报表如:

[0083] A、综合类:增值业务 KPI 统计、业务分省地区统计、业务各类成功率统计、业务流量统计、业务流量分布统计、业务量的用户分布统计等

[0084] B、终端类:系统注册终端型号数据排名、终端收发消息条数及成功率排名、终端收发 SP 消息条数排名及分布统计等

[0085] C、增值业务提供商类:SP 业务量及成功率统计、SP 上下行消息条数综合统计与排名等

[0086] 预定义报表的工作流程为:通过浏览器界面选择查询统计报表的条件,程序根据查询条件拼装针对 Cube 和关系型数据库的查询指令,该查询指令通过预定义的数据库/数据仓库访问接口传递给数据源,界面根据返回的数据集按照设定的呈现样式展现统计报表。所述的呈现样式包括 html、excel、csv、pdf 和 xml 等。

[0087] 本模块中,定制查询与报表的工作流程与预定义报表的流程区别在于:定制查询与报表的行列字段定义都是在界面中设定的,程序获取定制报表的行列字段定义进行语法分析,并进一步组装对于 Cube 和关系型数据库的查询指令。而预定义报表的查询主体中行列表定义、字段具体含义均预先定义为报表定义文件,查询过程只传递时间、地区或业务名称等少量参数。

[0088] 多维分析与预定义报表、定制报表的区别在于：

[0089] A、预定义和定制报表的查询条件和字段定义确定后，查询结果将以一张二维报表的方式进行展现。而多维分析中，界面展现所有 Cube 维度及度量值的要素。用户首先选定某些维度的取值，之后在界面上设置一个或多个维度作为分析列、一个或多个维度作为分析行、一个或多个度量值作为数据项。通过动态展开分析行及分析列可以获得在不同维度取值影响下的度量值数据，分析行及分析列维度在不同层次（level）上自动进行数据的聚合。

[0090] B、预定义和定制报表的数据源既包括关系型数据库，也包括数据仓库，而多维分析模块的数据源只基于数据仓库的 Cube。具体地，本模块的多维分析模块针对的数据源包括 ISDNCube（用户立方体）、KpiCube（KPI 立方体）和 SPCube（服务提供商立方体）等（如图 3 所示）。

[0091] 下面以一个典型的实例来描述业务数据多维分析的工作过程。

[0092] 原始需求：系统用户为提升服务质量，需要分析近两月以来彩信业务的所有终端到终端应用类型中，彩信终端手机收发彩信的 OT 话单条数、成功率变化趋势，并对成功率低于设定数值的情况，进一步分析其状态码（用于标识本条业务流程状态的编码，据此了解业务失败的原因等）的分布情况。

[0093] 工作过程：

[0094] 1、界面设定。

[0095] A、在用户界面中设定某些特定维度的具体取值或取值范围。包括设定终端类型维度取值为“彩信终端”、设定业务应用类型维度取值为“终端到终端”；

[0096] B、选定待分析的维度为 OT 话单类型维度、时间维度和状态码维度，同时设定该时间维度的起始和结束时间；

[0097] C、选定业务条数作为本次分析所关注的量化指标；

[0098] D、在界面上定义临时计算成员“成功率”及其计算公式。

[0099] 2、点击分析后，B/S 页面首先将界面的各种设定情况传递到统计分析的服务器端，在服务器端完成多维查询语句的封装，并送交数据仓库进行即时查询，查询结果数据不保存在服务器端，而是直接在用户界面的指定区域以多维报表或图表的方式进行呈现。

[0100] 3、如果需要分析其它指标，将重新进行步骤 1、2 的工作流程。

[0101] 综上，在完成消息类业务的数据采集、数据转换与装载后，系统具备了业务系统的多维数据模型，进而通过分析和呈现模块，提供用户预定义报表、定制查询和报表以及多维分析的方法和手段。通过这些分析和呈现手段，可为运营商和设备制造商提供更为灵活方便的数据统计方法及决策支持。

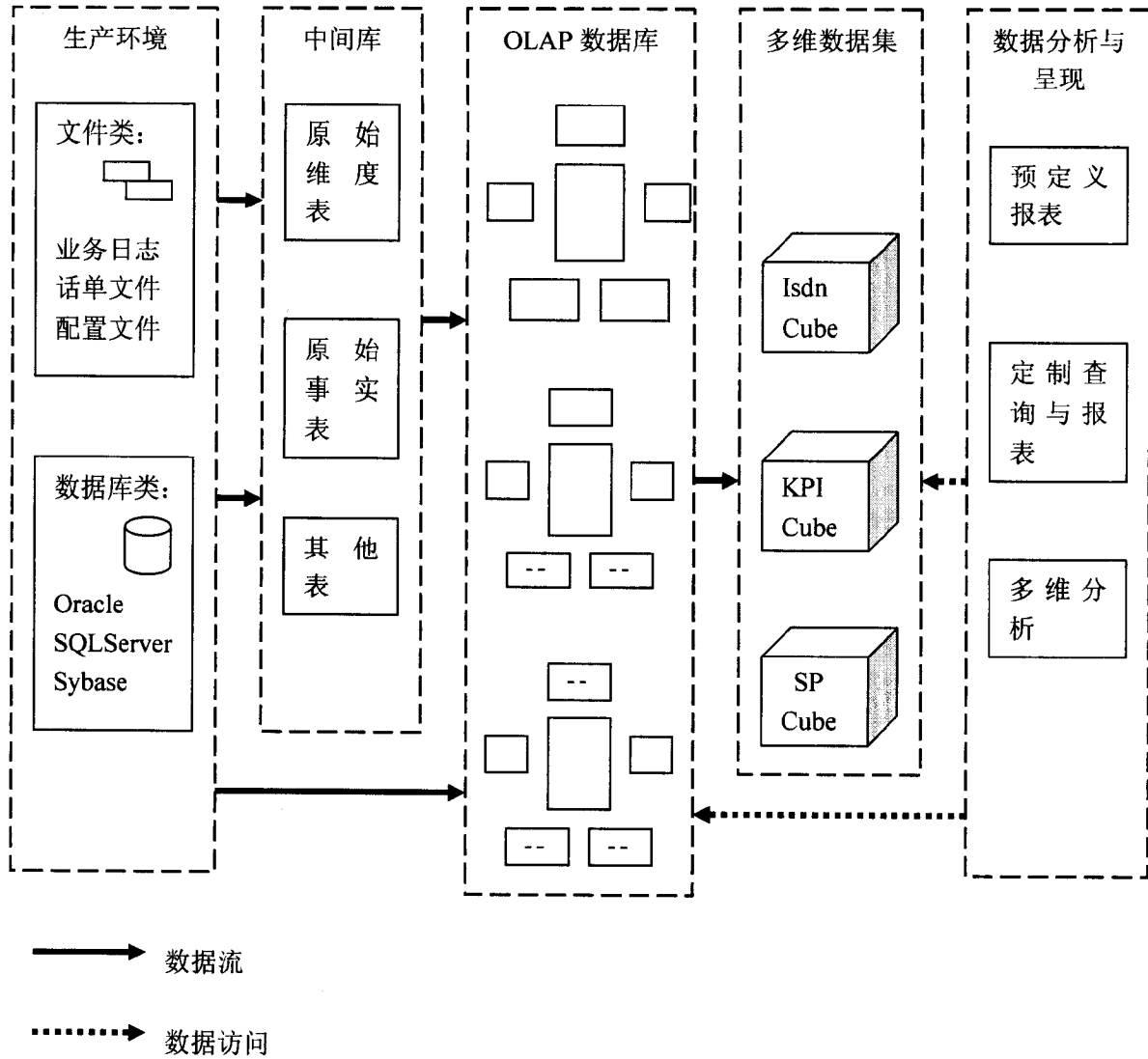


图 1

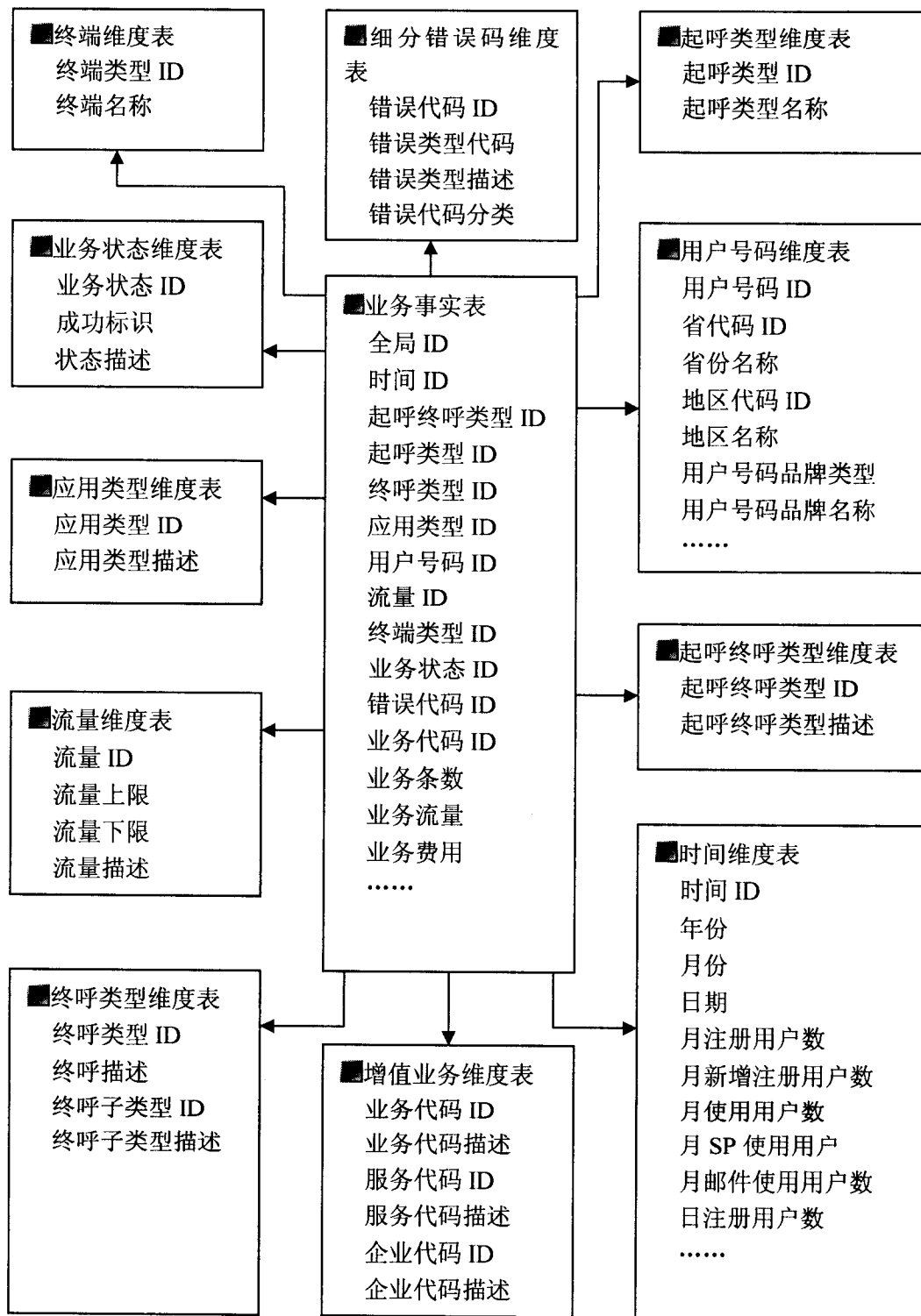


图 2

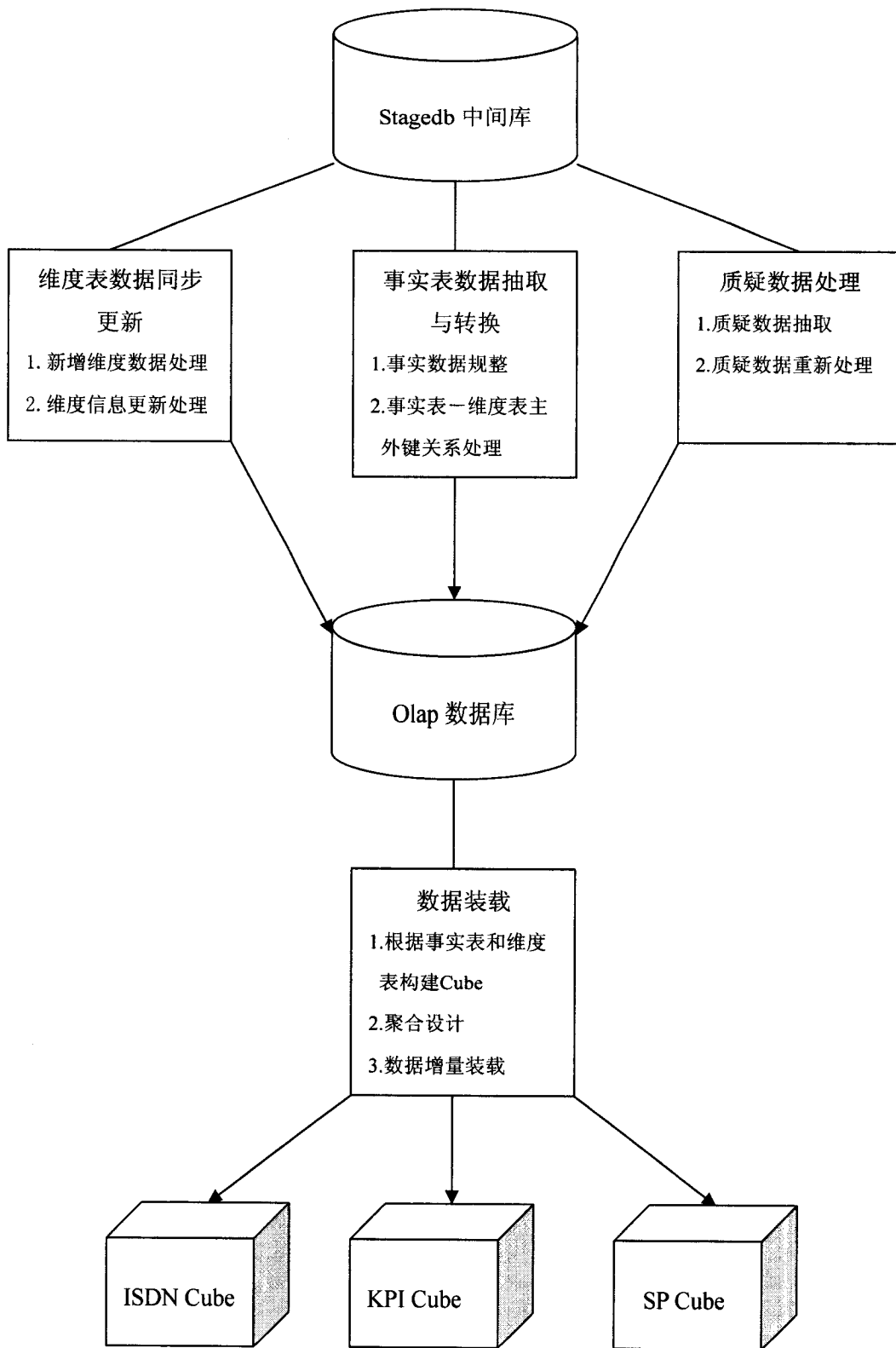


图 3