无线电监测数据仓库的构建与应用研究

田 斌[1] 朱亚磊[1] 张云春[2] 胡建陶[2] 张晨斌[2]
(云南省无线电监测中心信息科 昆明 650228)[1] (云南大学软件学院 昆明 650095)[2]

摘 要 对于无线电监测和相关海量数据的检测,单纯的存储和查询无法满足实际应用的需求。为满足无线电监测业务中的高层决策和智能监管业务需求,对业务系统中的数据进行预处理,设计并开发了无线电监测数据仓库。首先,为实现业务系统和信息化平台中源数据的采集,设计 ETL 规则;其次,设计维度、度量、级别,建立不明信号数据立方体;最后,实现多维数据模型的分析、预测和决策功能,从而增强不明信号相关业务功能。

关键词 数据仓库,不明信号,无线电监测,数据立方体
中图法分类号 TP399 **文献标识码** A

Research on Construction and Application of Radio Monitoring Data Warehouse

TIAN Bin[1] ZHU Ya-lei[1] ZHANG Yun-chun[2] HU Jian-tao[2] ZHANG Chen-bin[2]
(Information Department of Yunnan Radio Detection Center, Kunming 650228, China)[1]
(School of Software, Yunnan University, Kunming 650095, China)[2]

Abstract When faced with massive data on radio signal detection and surveillance, the simple storage and queries fail to meet the application requirements. In order to satisfy the requirements of higher level decision making in radio monitoring system and intelligent monitoring services, a radio monitoring data warehouse was designed based on the preprocessing of the data collected from existing radio monitoring system. First, ETL (Extract/Transform/Load) rules were defined to collect data from application systems and information platforms. Second, an unknown signal data cube was built after the design of dimensions, measures, and levels. Finally, the implementation of analysis, prediction and decision-making functions based on multidimensional data models was done. It helps to enhance the business functions related with unknown signals.

Keywords Data warehouse, Unknown signals, Radio monitoring, Data cube

1 引言

伴随着无线电管理系统的完善及日常业务的长期运转,无线电监测中心所需要处理的信息量迅速增长,因此迫切需要一种高效的解决方案来管理大量的信息。早期,无线电监测使用文件系统进行存储,将数据按一定的逻辑结构重组并保存在磁盘上。然而,实际应用的不断拓展对高速查询和数据共享的需求持续提高,而文件系统缺乏高度结构化的数据,不能满足需求的不断增长,很快便被数据库系统取而代之。数据库能实现数据的有效存储和快速查询,并提供安全机制、备份和恢复等高级功能。随着数据库系统的广泛流行,无线电监测中心的各职能部门分别建立了各自的业务系统。通过业务系统的数据采集和长期积累,各业务系统数据库中保存了大量业务有关的历史数据,数据规模不断扩大。因此,无线电监测中心的现有工作重点不再是简单的数据收集,而是从现有的历史数据中获取有意义的信息。

数据库管理系统在无线电监测系统中的应用普遍存在如下问题[1]:数据是非集成的,大量的数据独立存在于每个部门的业务系统中,相互之间不能共享,无法为高层决策提供必要的基础;数据库中的历史数据没有得到充分利用,大量的业务数据仅被进行存储和常规的查询操作,不能将数据用于支持业务系统的新规划和开发;各个部门之间的数据冗余,既增加成本,又阻碍了信息一体化建设;现有数据库系统支持联机事务处理(On-Line Transaction Processing, OLTP),主要执行日常的事务操作,对实时性要求较高,但数据量不大且不支持复杂的数据分析[2],而现有的监测系统对高层决策的需求日益提高。由于多个业务部门的存在,分布式数据库看似是一种完美的解决方案。分布式数据库是数据库技术与网络技术相结合的产物[3]。在分布式数据库中,数据在物理上是分布的,在逻辑上是集中的[4],但会造成部分数据的冗余。由于云南省无线电管理中心各业务系统存在自成体系的现象,异构的、海量的数据累积在数据库中,又缺乏专业的决策分析支持,难以将大量的数据信息汇总,不支持专业的决策和跨部门的查询,因此严重制约了单位的决策分析能力和信息化建设,对于利用这些结构复杂的海量数据来分析和辅助决策的需要日趋紧迫。而数据仓库是一种能提供良好数据分析和决策支持的解决方案。

数据仓库(Data Warehouse)是一个面向主题的、集成的、

本文受国家自然科学基金项目(61363021),云南省教育厅基金项目(2014Y013)资助。
田 斌(1982-),男,硕士,主要研究方向为无线电监测检测、信息化,E-mail:tb_sd@163.com;朱亚磊(1979-),男,主要研究方向为无线电设备检测、信息化;张云春(1981-),男,博士,讲师,主要研究方向为无线网络与网络安全,E-mail:yczhang@ynu.edu.cn(通信作者);胡建陶(1992-),女,硕士,主要研究方向为计算机网络;张晨斌(1994-),男,主要研究方向为无线电安全检测、Web 安全。

稳定的、反映时间变化的、用于支持管理决策的数据集合[1]。一方面，数据仓库面向分析型数据处理，用于支持高层决策；另一方面，数据仓库实现对多个异构数据源的集成，按照主题重组，保存不同粒度的数据，但保存的数据一般不能再被修改。数据仓库的特点正好能满足现有无线电监测业务系统中的需求。

基于上述分析，本文设计了适用于无线电监测中心的数据仓库体系结构。本文的创新点包括：1）完成对无线电监测各部分的业务分析，设计并实现了数据整合；2）基于维度和层次的定义，建模无线电监测数据仓库，并实现元数据抽取和管理；3）以“不明信号”为例，重点分析了不明信号数据仓库的构建，奠定了高层决策支持分析的基础。

本文第 2 节介绍数据仓库的有关研究成果和现状；第 3 节实现基于无线电监测中心的数据仓库体系结构，介绍源数据处理方法，并分析设计体系结构的关键；第 4 节以“不明信号”为例，重点介绍数据仓库设计的实现；最后总结全文。

2　国内外研究现状

数据仓库于 20 世纪 90 年代在国外兴起，促进了在线分析处理（On-Line Analytical Processing，OLAP）技术和数据挖掘（Data Mining）技术的发展[2-3]。由于业务分析和决策支持的迫切需要，出现了数据仓库，以满足新的分析处理环境对数据组织和存储的要求，它具有严格的投资回报率评估、整合了数据集市、增加了更多的分析以及动态数据仓库等特征。随着数据仓库的广泛应用，国外各大厂商（包括微软、IBM、Oracle、Sybase 等）均提出了自己的数据仓库解决方案，广泛应用于商业和研究领域。

Informix 于 1994 年发布了具备动态可扩展结构的数据库服务器，2000 年 IBM 收购 Informix 后，其发展更快。作为真正的可伸缩和可扩展数据库，它为用户的数据仓库提供强有力的支持，尤其当终端用户成倍增长时，效果显著。Informix 公司的数据仓库战略以体系结构的全新设计出发，确保了其解决方案的持续性。Oracle 是全球最大的数据库和数据仓库厂商，其产品强调满足 OLTP 性能和大数据量处理 OLAP（On-Line Analytical Processing）的双重需求。Sybase 也提供全套的数据仓库软件，采用“列存储”技术，具有并行查询及扩展能力强的优点，适用于事实分析。综合来看，IBM 的产品种类繁多，安装、配置和使用比较复杂，但提供仓库管理器，在数据仓库管理方面略有优势。Oracle 虽然没有提供类似的仓库管理器，但实现了图形化界面，可以进行快速设计，提供的许多新技术使其在数据仓库的性能方面占一定优势，适用于企业级的数据仓库解决方案。鉴于无线电监测中心各部门前期均采用了 Oracle 数据库，结合 Oracle 数据仓库的优势，本项目的开发使用 Oracle 数据仓库解决方案。

数据仓库的实现离不开 ETL（Extract-Transform-Load），现有的数据仓库解决方案厂商都设计了自己的 ETL 工具。不同的 ETL 工具各具优势：IBM 的 ETL 工具在抽取的速度和自动化程度上优于其他工具，但界面不够友好，处理复杂的数据源时存在问题；Oracle 的 ETL 工具提供数据提取、元数据管理功能，但在扩充性方面略有不足。

OLAP 是一种新兴的、支持决策的手段，以数据操作直观、分析灵活、可视化等特点在数据仓库的支持下迅速发展，为企业决策的支持提供良好解决方案。随着大数据时代[5]的到来，为实现数据的高效分析，更需要 Hadoop、Map Reduce[6]、内存数据库等新技术。这些新技术都丰富和推进了数据仓库技术的发展。

在国外企业中，数据仓库的应用已较为普遍，并呈现出应用较早、在电子化数据积累方面比较领先、有比较完善的管理和实施等特点[3-4]；在国内，许多企业也开始着手于数据仓库应用系统的建设。数据仓库在国内的应用主要包括：1）电信行业的应用：实现了对电信市场客户群的分布状况、消费特征、企业经营发展趋势分析等多方面的决策分析服务。2）销售行业的应用：对大量历史数据进行分析和挖掘，根据时间、地区等多个因素对销售量的影响，挖掘出有利的信息，对未来的市场进行预测；这对于提高同行业的竞争及降低成本都有着不容忽视的作用。3）金融业的应用：通过对业务过程中积累的关键数据进行采集，基于数据的分析，主要实现预测、预警和趋势分析等功能，以达到防范和化解风险的目的。数据仓库在金融业的应用主要集中在银行、证券、投资等领域，并出现了第四阶段的动态数据仓库[7]，应用领域更加广泛。

3　无线电监测数据仓库体系结构及关键技术

根据云南省无线电监测中心当前业务系统的实际情况，所构建的数据仓库体系结构如图 1 所示。从图 1 中可以看出，本项目设计的数据仓库主要包括 4 个部分，分别是数据整合层、数据服务层、应用分析层和信息展现层。其中数据整合层、数据服务层及应用分析层构成整个系统的数据中心。数据整合层主要实现数据的预处理、ETL 等业务；数据服务层侧重数据驱动下的数据仓库的迭代式开发；应用分析层侧重基于多维数据模型实现复杂的业务查询、数据挖掘等高级功能，为信息展现层提供基础。

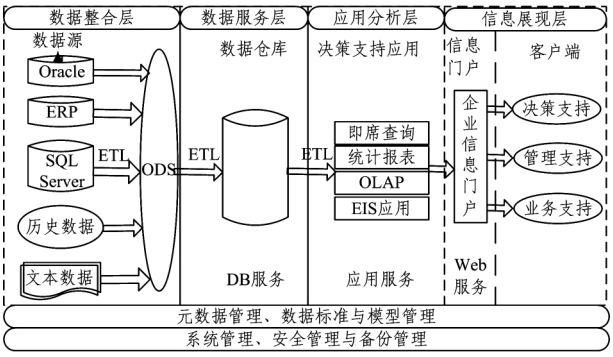

图 1　数据仓库体系结构图

3.1　数据整合

数据仓库的构建以数据驱动、数据为中心。因此，在使用现有业务系统构建数据仓库之前，首先要做好构建数据仓库前的软件和硬件资源准备工作。其中，最重要的任务是数据的来源、预处理和 ETL 工具开发。云南省无线电监测中心已经通过前期工作在各部门中部署了 Oracle 数据库系统，并开发了相应的 OLTP 应用程序[8]。本项目依托 Oracle 及其相关配件，对相关硬件、软件的安装等不再赘述。

（1）数据源。数据源是构建数据仓库系统的基础。现有的业务系统主要包括无线电监测、检测、台站和公共管理等多个业务系统。源数据大多来自于各个业务部门所安装和使用的 Oracle 数据库系统。此外，构建数据仓库的数据还包括许多外部数据，如文本数据和各种历史数据等。

（2）抽取、转换和加载。数据的抽取、转换和加载是本阶

段的主要任务，基于现有的业务中使用的数据库类型和各种需求，将操作型数据集成到数据仓库中。ETL 始终贯穿于数据仓库构建中，它是整个数据仓库的生命线，包括了数据清洗、整合、转换、加载等全部过程。

首先，将源数据进行清洗和整理，以满足数据仓库的要求，再放入数据准备区，即 ODS(Operational Data Store)层。ODS 层是整个系统中数据的统一入口，为数据仓库提供数据准备。ODS 存放从业务系统中直接抽取出来的数据，这些数据从数据结构、数据之间的逻辑结构上与业务系统保持一致，抽取主要关注数据抽取的接口、数据量大小、抽取方式等。ODS 也可以支持业务系统中一些报表的生成工作，分担一部分业务系统的查询压力。数据仓库主要保存经过汇总和压缩后的多粒度、多层次数据，因此 ODS 还可以完成数据仓库中不能完成的一些功能，例如：当查询的过程中需要用到最原始、粒度最小的细节数据时，可以把查询转移到 ODS 中来完成。

其次，设定抽取规则，制定抽取流程，将数据提取出来并进行必要的转换，再加载到主题数据库中。能够按照统一的规则集成并提高数据的价值是负责完成数据从数据源向目标数据仓库转化的过程，是实施数据仓库的重要步骤。无线电监测中心的数据虽然分布在各部门的子系统中，但是经过早期信息平台一体化的建设，基本实现了数据的融合。建设数据仓库时，主要涉及现有系统中数据抽取功能的实现。抽取后的数据不再按照传统的 E-R 图所定义的方式存储，而是重新组织，根据主题划分，以便进行数据的进一步分析。主题数据库是面向业务主题的数据组织存储，是针对各业务应用数据进行分析整理而设计的，不是各个业务应用数据的简单复制。对云南省无线电监测中心现有的业务进行分析，确定主题时要综合考虑。一个主题在数据仓库中即为一个数据集市，数据集市体现了整个业务系统某一方面的信息，多个数据集市构成了数据仓库整体。无线电监测业务中的主题包括：人员、监测、检测、台站管理、报表等。在确定了主题之后，分析主题所使用的度量，度量是数据抽取的指标，度量的选择至关重要；度量是构建数据仓库中的数据立方体，并进行立方体操作(如切片、分块、上卷、下钻和旋转)的重要因素。基于不同的度量可进行复杂关键性能指标(KPI)等的计算。确定度量之后，定义业务处理所涉及的粒度，考虑到该度量的汇总情况和不同维度下度量的聚合情况，采用“最小粒度原则”，即将度量的粒度设置到最小。粒度通常取决于度量属性本身的取值在层次上的不同划分，层次越多，划分的粒度就越多。

3.2 数据仓库建模与元数据管理

3.2.1 数据仓库建模

根据对云南省无线电监测中心各业务系统的分析，以需求分析和业务抽象得出的分析主题将源数据按照主题重新组织，划分维度，确定数据仓库的物理结构，同时对数据仓库的元数据按某种存储结构进行存储。实现时采用“维度建模技术”，以星型和雪花型模型来组织数据。

首先，对某一主题要确定用于分析处理的维度，确定维度的层次(Hierarchy)和级别(Level)。维度的层次是指该维度的所有级别，包括各级别的属性；维度的级别是指该维度下的成员。确定了主题及其维度后方可进行逻辑模型设计。由于星型模型能清晰地反映各种实体间的逻辑关系，并可在此基础上更好地组织检索和查询，因此可使用星型模型设计完善的数据仓库逻辑模型。星型模型有 3 个逻辑实体：指标实体、维度实体和详细类别实体。指标实体用户关心的实体，通常与主题相对应。维度实体用于限制用户的查询结果，与影响主题的属性对应。详细类别实体是维度内的一个单独层次。

其次，确定用于分析的数值型事实从而形成事实表，进一步考虑加载事实表。一般将指标实体转化为物理数据库表，也称事实表。维度实体通常也转化为维数据库表(称为维表)，它包括其每一层次的主码和对应的值。维表的关键字是该维度实体对应的详细类别实体的主码。维表和事实表通过维表的关键字进行关联。

事实数据表是数据仓库的核心，在进行 JOIN(联接)操作后将得到事实数据表，其记录条数一般较大。在保证数据完整性的基础上，为提高访问的效率和加快查询速度，设置复合主键和索引。事实数据表与维度表同时保存在数据仓库中，若前端需连接数据仓库进行查询，还需建立相关的中间汇总表或物化视图。

3.2.2 元数据管理

除正常的业务相关数据，系统中还存在另一种重要的数据，即元数据。元数据指关于数据的数据，主要用于定义数据的意义及系统各组成部分的关系，元数据由系统自动生成。元数据包括关键字、属性、数据描述、物理数据结构、源数据结构、映射及转换规则、综合算法、代码、缺省值、安全要求、变化及时限等。元数据用于建立、管理、维护和使用数据仓库。元数据管理是企业级数据仓库中的关键组件，贯穿于建立数据仓库的整个过程。在实际中，需要建立元数据库，用于保存元数据，服务于后续开发和维护。ORACLE 数据库中已有的 CWM(Common Warehouse Model)标准能兼容 ORACLE 数据仓库。

元数据通常被划分为技术元数据(Technical Metadata)和业务元数据(Business Metadata)。技术元数据存储关于数据仓库系统技术细节的元数据，包括：数据的逻辑模型和物理模型；数据仓库中的表名、字段名、关键字、索引及其相关属性；数据仓库数据与操作环境数据的对应关系和导入、过滤、校验的方法；进行 OLAP 分析所用的“维”和汇总数据的信息等。业务元数据包括保证用户能正确使用无线电管理的业务术语所表达的数据模型、对象名和属性名、访问数据的原则和数据的来源等。

3.3 数据挖掘与决策支持分析

数据仓库是进行决策分析的基础，建立数据仓库后可根据业务需要开发多种类型的应用[9]，尤其是支持系统分析和决策的工具，如各种报表工具即席查询工具、数据分析工具、数据挖掘工具及支持复杂分析操作的联机处理分析。传统数据库支持的 OLTP 支持高速的数据查询、更新和安全保证，而 OLAP 作为一种数据分析技术，被用于实现基于某种数据存储的数据分析功能。OLAP 对数据一般执行读操作，且一次访问大量数据，是面向主题的多维数据分析技术。

OLAP 工具完成的任务主要包括：给出数据的多维逻辑视图，视图独立于数据存储的具体形式；允许用户对数据进行交互式查询和数据分析(交互式操作有多种方法，包括钻取、切片和切块等)；检索并显示多维表格、图表和图形中的数据，便于坐标轴位置的变换；以较快的速度响应查询，缩短用户的访问时间。此外，利用 Oracle 提供的挖掘工具 Oracle Data Mining(ODM)，实现关联、聚类、分类、预测、时序模式和偏差分析等 6 种信息挖掘方法。

3.4　信息展示

信息展示主要实现可视化功能，按照用户的分析需求，把数据仓库系统中的信息和分析结果提供给最终用户。可视化使用定制报表、随即查询、多维分析和数据挖掘等方法和技术进行数据拓展。

针对同一主题的数据，OLAP 可以从不同的角度对其进行展示，根据终端用户的需要，随意组合展示的角度和展示的方式。其次，OLAP 既能提供数字报表展示，还能提供强大的图形化界面展示功能；可以对数字报表以柱状图、饼图、折线图等图形直观地展现给用户，支持对用户关心的图形区域做进一步细化展示的功能。利用上述工具可以方便地把无线电监测和检测有关的信息以各种图表方式展现出来，帮助无线电管理中心做出正确的高层决策。

4　基于不明信号的数据仓库构建

通过对云南省无线电监测中心各业务数据的分析，本项目以不明信号为主题，设计了数据仓库模型的实现方案。不明信号指技术人员在日常监测等工作中发现的、信号频率未在无线电管理部门登记注册的、在所辖范围内无线电管理机构频率台站数据库中无记录的无线电信号。监测到信号后，首先将其与已保存的合法信号集合进行比对，若发现异常则及时查明信号源的性质、方位、发送设备、使用单位等信息，并发出相应的提示信息。随后将其录入到数据库中，为后期频谱管理提供技术依据和保障。

针对不明信号的管理，现有的功能局限于不明信号检测和信息存储，是一种被动的行为模式，无法适用于新环境下的无线信号管理。在现有的业务系统中，随着信息一体化平台的建设，对不明信号的相关数据应进行有效的挖掘和分析，从而能准确地标注不明信号并及时地进行排查，这对于提高无线电监测人员的工作效率至关重要。

通过对业务系统的分析，不明信号的管理包括：频段扫描、监测统计报表、站点统计、地区分布统计等功能。建立不明信号数据仓库，可以实现对不明信号数据的复杂查询，并为高层业务决策提供依据。所建立的不明信号数据仓库应实现的功能包括：1）通过数据立方体的构建和分析，判断不明信号出现的时间规律、内容、口音等内容所显示出的特征；2）基于信号的波形对不明信号相关业务类型进行分类；3）固定监测站和移动监测站部署模式的分析；4）不明信号种类的分析与新检测信号类型的预测，不明信号的常见种类包括[10]：非法发射、合法未备案发射、遗漏登记的发射、非本区域的发射、非无线电发射、杂散发射、公共频段和业务频段发射等；针对每一种类型，通过数据仓库技术结合分类算法，提取分类特征，对新检测到的信号进行特征匹配，预测该信号的类型；5）不明信号的标注和排查，将标注的结果写入数据库之前必须保证其准确度；排查后要进行发射源的查找。

4.1　不明信号星型模型的设计与实现

4.1.1　影响不明信号的因素

以不明信号为主题，构建数据仓库时所需的数据包括：监测值班人员（值班的频率、值班开始时间和结束时间）、监测执行（监测站的状态、监测站的类型）、监测任务（监测任务类型、不明信号的条件）、设备信息（设备的健康状况、设备的类型）、时间（粒度有年、月和日）。此外还包括创建上述数据时生成的元数据。

4.1.2　为不明信号构建的星型（雪花）模型

针对现有的业务数据库，首先判断建立数据仓库时所要抽取的数据（包括数据表、元数据等）。根据业务分析得到不明信号的影响因素，以不明信号为主题建立的星型模型如图 2 所示。

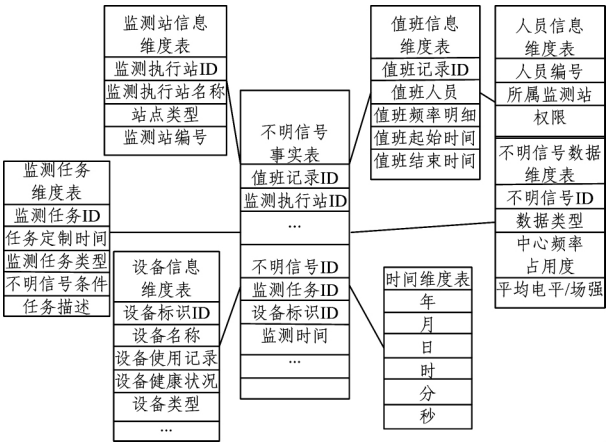

图 2　不明信号星型模型图

从图 2 中可以看出，不明信号星型模型的核心是不明信号事实表，通过表中的属性与各维表相关联。维表包括：人员、时间、监测站、设备、监测任务等。

4.2　数据抽取规则的设计与实现

数据抽取时应保证数据的完整性和一致性。实现时采用 OWB（Oracle Warehouse Builder）组件来完成数据的抽取。首先，以星型模型明确需要抽取的数据，将数据分类，将抽取的字段放到相应的分类中，合并相似字段。其次，对于编码类型、编码方案、数据粒度等的不一致问题，按照相同的规则抽取到目标数据库中。以时间维为例，在不同的业务数据库中，该维度存在月、日、年等不同粒度的差别。OWB 中提供多种抽取操作，可以根据条件对数据进行联合、拆分和过滤，这些操作都进行了打包，无需额外的开发就可以使用这些功能。最后，按照设计的抽取规则，利用 Oracle 的 ETL 工具，可以完成数据的抽取过程，并且可以验证抽取数据的完整性、正确性、一致性。

基于现有的大数据处理平台，需要对已有的 ETL 工作流程进行改进和优化。对 Oracle 的现有 ETL 工具的改进包括：1）将 ETL 细化，对转换过程进行最细粒度的管理；2）任务配置文件。通过编写任务配置文件，将任务规则设计中的配置文件出现的标签和插件的引用同整理框架相关联。

4.2.1　抽取的时间

在系统的实际运行过程中，抽取策略也是一个关键的设计过程，包括抽取时间、抽取周期等。抽取时间应根据系统处理数据的时间来定，无线电监测的各个业务系统的工作时间不同，应该为各个系统设置不同的数据抽取时间。例如监测系统，大量的业务数据会在白天的频繁业务操作下产生，所以抽取时间应该设定为合适的晚间时段。

4.2.2　抽取的周期

对于监测站、人员和设备等实时分析要求不高的数据，可设计固定的抽取周期，一般为一周。对于不明信号数据，可以将不明信号出现日期作为抽取周期，对阶段性数据进行分析，以便实现不明信号在时间、区域、设备类型、关联性等方面的高层分析。其他数据的抽取周期依据更新频率可灵活设置为每天、每周或每月。

4.3 多维数据模型的构建

数据仓库的构建以数据为驱动，通过迭代式的方式实现。在数据源和相应的数据抽取规则确定之后，构建数据仓库的核心内容包括以下步骤。

4.3.1 设计维度

在定义维度时，首先创建维度的属性，在此基础上创建级别，分析各级别的层次关系，设置级别的属性。对于不明信号相关的时间维，其级别为年、月、日等。以此级别对不同粒度的数据进行加载。在不明信号事实表中设计了监测站维度，通过维度表详细给出监测站有关的属性和信息。此外还有时间维、不明信号维等。

4.3.2 数据立方体的设计

立方体的构建是数据仓库能够支持复杂查询和高层决策的关键。根据上一步骤中建立的“维”，选择相应的“维”作为立方体的基础，并定义“度量(Measures)”，从而能够支持数据立方体的多种操作。以监测站、时间和不明信号量 3 个维度定义的数据立方体如图 3 所示。

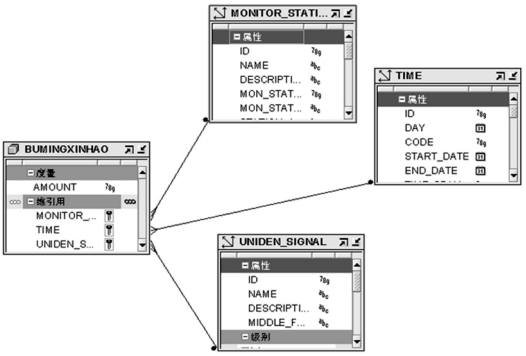

图 3 不明信号数据立方体

4.3.3 设计映射

映射描述了从数据源抽取、转换及加载数据的一系列操作，并提供数据流和针对数据流操作的可视化展示。映射的关键在于“运算符”的设计，使用运算符表示数据流中的源和目标，也可以使用运算符定义从源到目标的数据转换。以时间维度的映射为例，其映射如图 4 所示。

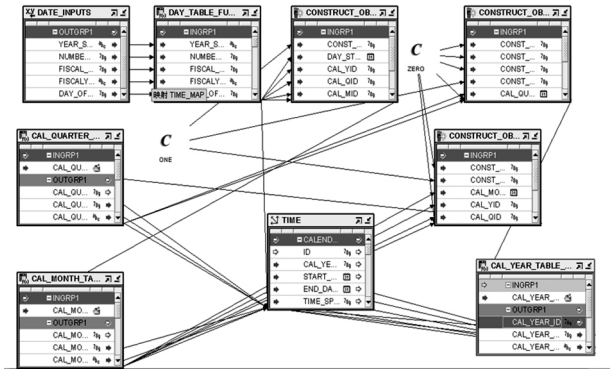

图 4 时间维度映射

从时间维出发，经过映射，选择相应的运算符才能实现对数据的聚合。以日期相关数据为基础，选择运算符，从具体的日期聚合到月份、再到季度、再到年份的运算，都可以使用图 4 所示的 C(count)运算符实现。对于一些简单的维，则无需使用复杂的运算符，可以直接映射得到，如不明信号维。

4.3.4 设计工作流

工作流的作用是将映射串联起来，确定映射运行的先后顺序。映射可以串行执行，也可并行执行。

4.3.5 设计计划

不同的“工作流模块”和“目标数据模块”需要设计不同的“计划”。

4.3.6 配置、部署、运行

配置的目的是将上面步骤中创建的进程流和计划进行关联，即保证“工作流”按照“计划”执行。

4.4 多维数据模型的使用和决策分析支持

以数据仓库中建立的多维数据模型为基础，可以实现比常规的数据库更复杂的查询。对多维数据进行查询能获得对数据更深入的了解，同时也能够作为数据挖掘、大数据分析、商业智能等高级应用的基础。本项目设计的不明信号数据仓库可以进行多维数据的查询和展示。以不明信号出现的日期进行的多维数据分析为例，其查询的结果如图 5 所示。

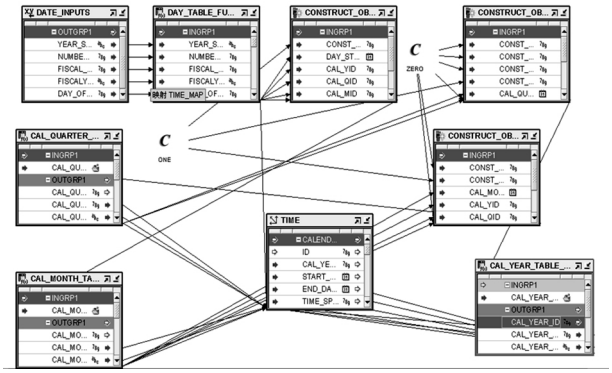

图 5 不明信号在时间维度上的查询结果展示

复杂的查询可以同时支持多个维度，以时间维和地区维组合查询为例，查询结果如图 6 所示。高级的查询功能可以基于多维数据，但是超过三维以上的结果在展示方面还有待改进。对数据仓库的开发还应包括数据挖掘、商业智能分析等，这些将在下一步工作中深入展开。

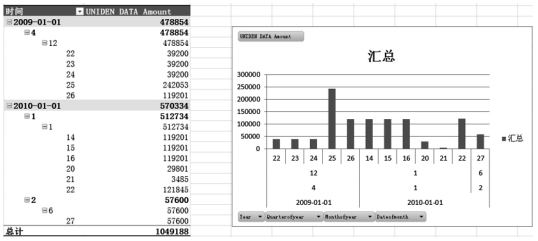

图 6 按时间和地区维度展示数据

现有系统基本实现了所有复杂查询和综合分析的功能，但以不明信号数据立方体为基础，仍然需要开发高级的功能，包括：新不明信号出现的时间和频率预测、不明信号监控的空间分布成因分析、新不明信号种类的识别和分类等。

结束语 本文设计并构建了基于 Oracle 数据库和数据仓库工具的云南省无线电监测数据仓库，实现了以不明信号为主题的数据仓库开发和应用。该数据仓库为不明信号有关的决策和数据挖掘奠定了基础。未来将在以下方面进行改进和完善。首先，实现基于数据仓库的数据挖掘和智能化，包括聚类分析、模式挖掘、不明信号出现的预测等。在使之与现有的研究相结合，实现对不明信号的全覆盖范围智能化监测和检测。其次，在数据的组织查询方面，设计更有效的计算方法以提高查询效率，如多维索引树的构建、云环境的搭建等。在

(下转第 542 页)

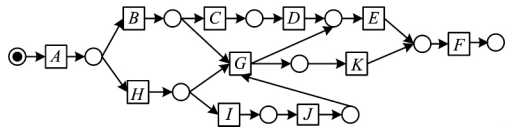

图 3 调整模型 M_1

对于调整后的模型，计算模型 M_1 与日志 $AHGIJKF(486)$ 的行为紧密度 $\xi \approx 0.866$；显然，对于日志 $AHGIJKF(486)$，模型 M_1 还不是优化的模型。

上述通过模型 M_0 与日志的行为紧密度进行调整，得到了优化的模型 M_1，但这不是最优的模型，因此还需要继续调整模型 M_1，调整后的模型为带配置信息的模型 M_2，如图 4 所示。

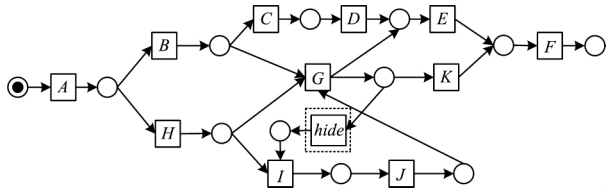

图 4 模型 M_2

计算模型 M_2 与日志 $AHGIJKF(486)$ 的行为紧密度 $\xi \approx 0.94 > 0.90$，根据算法可知模型 M_2 是最终优化后的模型。

结束语 本文基于模型与日志的紧密度对业务流程模型进行优化分析，首先通过频数较高的日志建立初始模型，由于不是全部的日志，因此初步建立的模型可能会存在差异，需要对模型进行优化分析。本文通过计算模型与剩余的与模型存在行为差异的日志的行为紧密度来优化业务流程模型，并通过配置变迁进一步优化模型，最后用一个实例说明了业务流程优化方法是可行的。

虽然对业务流程优化得到了较符合日志的模型，但还是存在一些问题，如何找到添加配置变迁的位置将是下一步工作的重点。

参 考 文 献

[1] WEBER B, REICHERT M, MENDING J, et al. Refactoring large process model repositories [J]. Computers in Industry, 2011, 62(5): 467-486.

[2] VAN DER ALST W M P, DUMAS M, GOTTSCHALK F, et al. Preserving correctness during business process model configuration[J]. Formal Aspects of Computing, 2010, 22(3-4): 459-482.

[3] VAN DER ALST W M P, LOHMANN N, ROSA M L. Ensuring Correctness during Process Configuration via Partner Synthesis[J]. Information Systems, 2012, 37(6): 574-592.

[4] REIJERS H A, MENDING J, DIJKMAN R M. Human and automatic modularization of process models to enhance their comprehension[J]. Information Systems, 2011, 36(5): 881-897.

[5] YOUSFI A, SAIDI R, DEY A K. Variability patterns for business processes in BPMN[M]. Springer-verlag New York, Inc. 2016.

[6] BOURNE S, SZABO C, SHENG Q Z. Managing Configurable Business Process as a Service to Satisfy Client Transactional Requirements[C]// 2015 IEEE International Conference on Services Computing (SCC). IEEE, 2015: 154-161.

[7] GRUHN V, LAUE R. Reducing the cognitive complexity of business process models[C]// The 8th IEEE International Conference on Cognitive Informatics (ICCI). 2009: 339-345.

[8] JIMÉNEZ-RAMÍREZ A, WEBER B, BARBA I, et al. Generating optimized configurable business process models in scenarios subject to uncertainty[J]. Information and Software Technology, 2015, 57: 571-594.

[9] HUANG Y, FENG Z. A Validation Method of Configurable Business Processes Based on Data-Flow[C]// Service-Oriented Computing-ICSOC 2014 Workshops. Springer International Publishing, 2015: 323-335.

[10] GRUHN V, LAU E R. Reducing the cognitive complexity of business process models [C]// The 8th IEEE International Conference on Cognitive Informatics (ICCI). 2009: 339-345.

[11] 吴哲辉. Petri 网理论[M]. 北京：机械工业出版社，2006：6-42.

[12] WEIDLICH M, POLYVYANYY A, DESAI N, et al. Process compliance measurement based on behavioral profiles[C]// International Conference on Advanced Information Systems Engineering. Springer Berlin Heidelberg, 2010: 499-514.

[13] KUNZE M, WEIDLICH M, WESKE M. Querying process models by behavior inclusion [J]. Software & Systems Modeling, 2015, 14(3): 1105-1125.

[14] WEIDLICH M, POLYVYANYY A, DESAI N, et al. Process compliance measurement based on behavioural profiles[C]// International Conference on Advanced Information Systems Engineering. Springer Berlin Heidelberg, 2010: 499-514.

(上接第 514 页)

数据的展示方面，也可以结合 B/S 模式的 OLAP 分析，提供更直观的展示效果。最后，对于无线电监测，不明信号的发现是一个至关重要的问题，现阶段的数据仓库系统在考虑的因素方面仍存在不足，在接下来的开发过程中需要进一步地对其进行完善，更好地为无线电监测中的决策问题提供支持。

参 考 文 献

[1] INMON WH. 数据仓库[M]. 王志海，译. 北京：机械工业出版社，2006.

[2] 周渝霞，刘道践，郝玉清. 基于 Oracle 的 OLTP 与 OLAP 数据库设计及实现[J]. 电脑编程技巧与维护，2012(10)：29-31.

[3] 左翔，姜文彪. 分布式数据库系统的设计与优化[J]. 赤峰学院学报，2012，28(10)：20-21.

[4] 顾玮. 分布式数据库技术在教务管理系统中的应用[J]. 办公自动化，2009(24)：50-52.

[5] 陈世敏. 大数据分析与高速数据更新[J]. 计算机研究与发展，2015，52(2)：333-342.

[6] 王有为，王伟平，孟丹. 基于统计方法的 Hive 数据仓库查询优化实现[J]. 计算机研究与发展，2015，52(6)：1452-1462.

[7] 武彤，谭光炜. 基于索引视图实现动态数据仓库的实时数据加载[J]. 计算机科学，2016，49(6A)：493-496.

[8] 苗晟，董亮，何丽波，等. 基于小型监测站的无线电监测系统构建[J]. 电子测量技术，2014，37(5)：132-135.

[9] 杜远宗，张金刚，夏堃，等. ETL 工具在建设数据仓库中的应用[J]. 中国科技信息，2005(8)：6.

[10] 崔忙良，张文彬. 浅析不明信号的有效标注与排查[J]. 中国无线电，2005(1)：46-47.