

基于EAM的水环境数据平台的设计与实现

陈 岩¹, 赵翠平¹, 马鹏程², 王 东¹

(1.环境保护部环境规划院, 北京 100012; 2.北京中创聚智信息技术有限公司, 北京 100012)

摘要: 通过研究水环境管理中的数据库或数据平台的规范建设模式, 提出了基于EAM的水环境数据平台建设方法, 采取EAM方法设计了分层数据库, 引入了数据集市对业务模板分类管理, 构建了多数据库兼容技术和算法生成模块, 综合建设出一套适合大数据管理要求的水环境数据平台。实践表明, 这种方法构建的数据平台, 能有效解决目前多系统、多数据、应用需求动态变化、以及数据安全的问题, 有利于提升水环境大数据建设的基础。

关键词: 水环境; 大数据; 数据平台; EAM 数据集市

中图分类号: X853

文献标志码: A

DOI:10.16803/j.cnki.issn.1004-6216.2016.05.007

Design and Implementation of Water Environment Data Platform Based on EAM

Chen Yan¹, Zhao Cuiping¹, Ma Pengcheng², Wang Dong¹

(1.Chinese Academy for Environmental Planning, MEP, Beijing 100012, China;

2.Sino-intelligence Information Technology Co., Ltd., Beijing 100012, China)

Abstract: In this paper, specification construction modes of the database and data platform of water environment management was studied and methods for constructing water environment data platform based on EAM was proposed. EAM was used to design a hierarchical database and data mart was introduced to classify and manage the business templates. Multi-database compatible technology and algorithm module was established. Therefore, a set of water environment data platform suitable for meeting the requirements of big data management was constructed in an integrative way. The practice showed that the data platform constructed by application of this method could effectively solve many issues concerning multisystem, multiple data, dynamic changes of application demands, data security, etc., helpful for improvement of big data construction of the water environment.

Keywords: Water Environment; Big Data; Data Platform; EAM Data Mart

CLC number: X853

党的十八届五中全会在《中共中央关于制定国民经济和社会发展第十三个五年规划的建议》中提出实施“互联网+”行动和国家大数据战略^[1]。我国环境管理模式正处于从以环境污染控制为目标导向向环境质量改善为目标导向的转型时期^[2-4], 随着水污染防治行动计划的颁布实施^[5], 以及国家信息化建设工程和环保领域大数据的建设启动^[6-7], 水环境管理的信息化建设将全面开展。

水环境综合管理涉及自然、社会经济、污染源、水质等众多要素, 信息庞杂, 数据庞大。当

前, 水环境信息系统众多^[6,8], 但数据库(数据平台)建设缺乏标准和深入研究, 数据的独立处理机制建设缺乏, 数据管理效率不高。随着数字化建设的深入, 越来越多的应用系统和数据信息上线^[9], 格式更加庞杂、数据更加庞大, 数据库的规范化建设成为水环境管理信息化建设的一项重要内容, 是流域信息化管理的重要基础^[10]。数据仓库与数据挖掘属于应用驱动, 这要求在具体应用领域, 特别是水环境管理中, 应该统一格式与建设规范, 并且度身定制专属的数据仓库与数据挖掘方案^[11]。

收稿日期: 2015-10-19

作者简介: 陈 岩 (1981-), 男, 硕士。研究方向: 水环境模拟与信息化。

通信作者: 王 东 (1973-), 男, 研究员。研究方向: 水环境污染防治规划、政策技术研究。E-mail:wangdong@caep.org.cn

数据库模型中，传统的关系表模型将所有的基础信息，即属性组成为一张表的字段，这样的优点在于，数据的读取简单，但是出现业务的变更时，若对属性进行相应的增加和删除，就需要对相应的数据表进行更改，操作繁琐，有一定的风险。EAV (Entity - Attribute - Value, 实体-属性-值) 数据库模型，优点是可以动态为数据模型增加或移除属性，其将模型的属性以记录的形式存放在数据表中，需要新增和移除属性时，仅需要删除数据表中相应的记录即可，而无需修改数据库结构，便于业务的扩展，但是同一记录的不同数值要按照类型分散到多个表中，查询记录需要关联多张表，并且要执行大量的列转行操作；添加、修改、删除数据需要关联Value表，由于操作需要大量的关联，所以执行效率非常低，业务系统开发和维护难度高。

文章提出一种新的数据库模型 - EAM (Entity - Attribute - Model, 实体-属性-模板) 模型，集合了传统的关系表模型和EAV模型的优点，以基础属性表、逻辑表、实体表3类构建分层数据库，将基础数据、算法模板、业务数据分层保存和管理，避免基础数据和业务数据的交叉影响，提升了数据安全性和数据库的管理效率；进一步构建了数据集市，将数据关系、算法模板集成管理、分类保存，实现了业务功能的动态保存和实时调整，提高了系统开发效率；开发了数据库兼容技术和算法设置模块，促进了数据库的多元融合，提升了系统管理的易用程度和管理效率。综合上述理念构建了水环境数据平台，实现数据存储高效、优质、有用，促进数据库建设的规范统一。

1 水环境信息系统与数据平台建设现状

水环境管理系统大量上线，包括环统、监测、监察、监控、环评、排污收费等众多系统。但是，系统均独立设计，之间通信尚未建立，数据共享困难。存在如下问题。

基本功能重复建设。各种基本功能，如统计分析模块、水质评价模块、用户管理模块等，在不同水环境管理系统中均有需求，不同系统中这

些基本功能均有建设，造成了资源浪费，并且数据格式不同，数据整合困难，数据库和应用系统维护成本过高。

系统复用性差，更新或开发成本高。系统开发尤其是数据库开发缺乏统一标准，各系统技术不同，开发兼容性差，可复用的功能少，因为要与旧系统程序与数据衔接，新的开发和改进耗时耗力，成本较高。

数据分散，不利于综合分析。系统主要针对专项需求开发，系统服务目的单一，各系统多是独立数据库，数据库格式针对性强，数据分散，数据差异大，不利于数据综合管理与分析，数据挖掘的难度大。

底层数据库设计与建设不足。众多系统是综合建设系统，底层数据库缺乏独立设计，数据库建设不足，面临环境管理需求不断更新，如考核办法、评分办法等的更新，难以动态跟踪修改，数据库更新跟不上管理需求的变化，系统整体修改或者更新难度大。

缺乏数据分类处理机制，难应对大数据处理要求。多数系统的基础元数据、业务计算、模块管理数据共同运行，缺乏数据分类处理机制，不同类数据表混合处理。信息化背景下，水环境数据从社会经济、环境质量、污染源、环境管理等全方位增长，数据多源异构，不断增大的数据处理量，将使系统运行速度越来越慢，难以应对大数据管理要求。

2 水环境数据平台的核心设计

水环境数据管理平台需要具备应对业务动态变化的能力。如何同时降低业务变化带来的数据库风险并提升数据使用效率是水环境信息化建设的关键。应对业务变化，一般数据管理平台需要修改或者创建新的数据表对象，这需要修改和开发相应的系统模块，甚至会导致整个系统的重构，具有很大的风险。而实施数据管理平台的分层设计可以解决这类问题。

2.1 基于EAM的数据平台分层设计

基于EAM (Entity Attribute Model) 思路，以

分层的办法,将原数据(基础数据)与逻辑数据(算法模板、业务数据)分别保存,使数据库能动态应对业务变化,同时又能维护原数据不变。不同于传统数据库设计,EAM将数据库分成属性、模板(逻辑表)、实体3层。属性是指可以量化或者被记录的特征,也即数据的基础信息值,如温度、湿度、污染物浓度、名称、位置等信息;每个属性的数据库格式固定,如数据类型、长度、是否非空、注释等。模板,即逻辑表,是指通过一定算法实现的属性集合,如水质数据评价模板,包含了断面、区域、时间、污染物、是否达标等信息;它代表了一个业务集合,目的是将业务信息整合,形成信息集成,提高计算效率,同时提高在应对业务变更的信息灵活性。实体,实体是指数据库里的数据表,每个实体表都对应一个模板,它是模板的持久化结果,数据最终保存在实体中。

通过这3层设计,将数据库多层化,采取属

性集最大限度的保留原始数据,采取逻辑表关联业务,使属性数据层不直接与业务发生关联,形成业务与原始数据的分别存储,应对业务要求持续变化的情况,不需要底层数据的整理与修改,保留了底层数据的原始性。

应对业务变化时,这种设计不需修改原始数据表结构,只需根据原始模板和业务去创建新的模板,将新模板创建到数据库中,支持新业务的开发和运行。原始模板对应的数据表依然保留,最大化的保留了历史数据,同时又能支撑原始版本的模块正常运行。并且,在不同模板间,原始版本数据可以按照规范较容易的同步到新模板数据表中,当新业务模块开发完毕后,就可以进行平稳的切换,保证业务的完整性。这增强了数据适应性和安全性,容易适应业务系统中数据模型频繁变化。这种分层次的设计方案,提升了数据库的业务整合性,为整个数据平台和业务平台的扩展奠定了基础。EAM基础设计见图1。

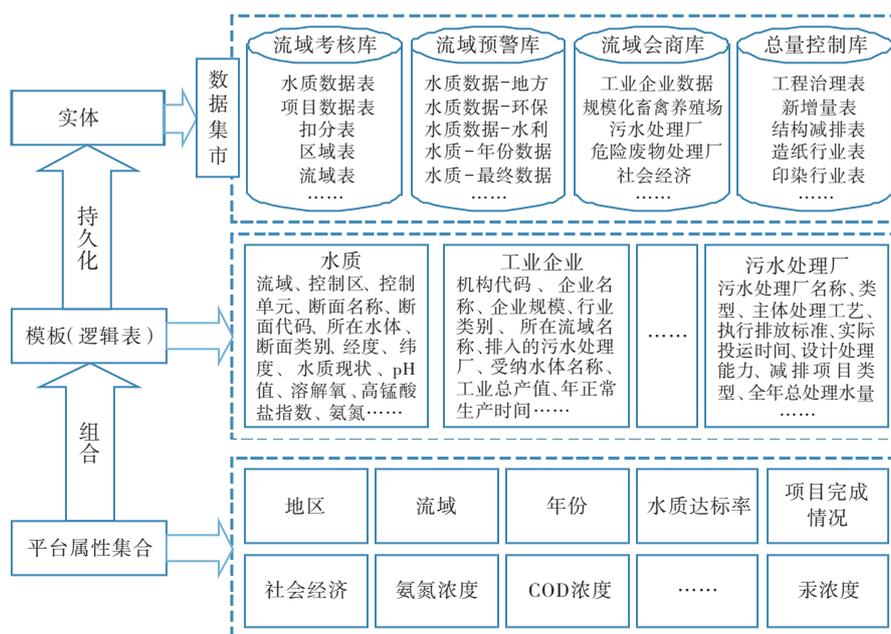


图1 EAM基础设计

2.2 采取数据集市实现分类动态管理

EAM中,“模板”由一个或多个“属性”组成,“模板”在数据库中持久化变成“实体”,支持业务系统。不同业务需求生成若干实体,形成实体集合,但这些集合,需要进一步优化,以方便统一管理,提高数据库管理效率。在EAM的基础上,将实体对应不同的业务要求进行归类,

将这些模板、逻辑表持久化,分类放置不同实体库中,形成数据集市,见图2。

数据集市是由多个逻辑库(模板)组合而成,是多个实体分类别(业务或者逻辑库)的组合。数据集中,针对新增业务,需要构建或完善的新实体和数据,通过直接操作模板(逻辑库)实现,提高了数据管理与业务运营的效率。

数据集市中的一类实体对应一类具体的业务，是一个业务逻辑库的群集，保持了数据集市与底层原始数据库之间的松散耦合，保持了业务更改便捷的同时又不会影响到底层数据安全，实现了在水环境管理业务上的灵活性和安全性。

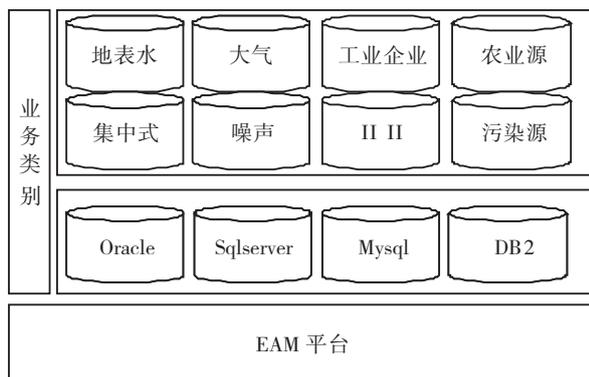


图2 数据集市设计

3 水环境数据平台的实现

平台数据中心的总体思路是在EAM的基础上，以数据集市的理念，将所有业务系统的数据库和对象映射入数据平台。平台建设支持多种主流数据库，包括：oracle、sqlserver、mysql等。平台包含实体与虚拟的结构型关系数据，是一种大型数据库，里面包含着水环境管理业务与系统数据，可以根据业务类型进行分类管理，根据EAM和数据集市的构建规则，将数据归类、整合，保存到不同的数据表 and 数据库中，整体形成数据管理平台。

平台建设业务模块对数据统一管理，实施批量操作、统计分析和自定义结果展示，包括查询、修改、删除、添加、多维度统计分析、数据导入、数据导出等数据操作，并可以直接通过平台在数据库创建新的数据对象，如表、视图等。

在EAM与业务模块的基础上，构建数据库兼容技术模块，增加数据库兼容性，进一步增加通用算法模块，将数据库操作和逻辑计算整合，实现数据库动态修改、操作、统计分析等。根据不同时段的业务需求，将算法封装，发布成系统内服务，供各业务系统调用，快速满足新的业务需求。业务人员业务根据动态需求，动态调整，动态调整数据平台算法模块，直接调整数据系统，

满足业务需求。

平台实现了对所有业务系统的底层数据完全控制，又能动态满足对外需求提供各种计算服务。以Webservice、SOA为基础，对算法封装成WEB服务，统一将服务交给ESB（企业级服务总线）托管，为各种业务系统提供服务。平台总体设计见图3。

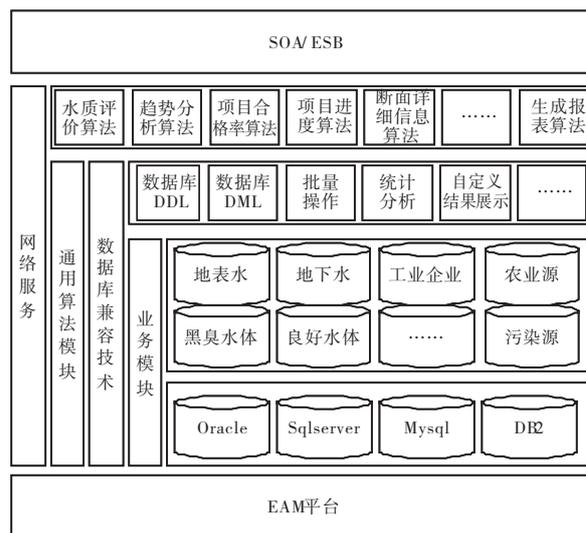


图3 数据平台设计

4 常用技术改进

4.1 数据库兼容性优化

平台建立了数据库兼容性算法。数据库兼容主要有两个问题需要解决：语法问题、数据跨域问题。Antlr是一个基于Java开发的功能强大的语言识别工具，Antlr以其简介的语法和高速的运行效率是这类工具中较高级的一种。采用Antlr开发语言重新定义SQL语法，分别针对oracle、mysql、sqlserver等主流数据库的语法进行整理，统一成一套标准的SQL规范。

支持如下几种类型的SQL关键字以及命令语句：CreateTable/AlterTable/DropTable/ CreateIndex/ DropIndex/Select/Insert/Delete/Update；支持子查询、Join、Union等高级的SQL特性；支持日期（包括取当前日期、从日期中提取任意部分、计算日期差异、日期前后推算等）、数学（包括取绝对值、取PI值、四舍五入、对数计算、随机数等）、字符串（包括取子字符串、取字符串长度、字符串截断、大小写转换等）、基本数据处

理(包括数字字符串互转、日期转字符串、非空判断等)等函数。支持的数据类型包括:“int”、“integer”、“tinyint”、“smallint”、“bit”、“boolean”、“varchar”、“decimal”、“numeric”、“real”、“float”、“double”、“datetime”、“date”、“timestamp”、“time”、“blob”、“tinyblob”、“longblob”、“binary”、“varbinary”、“longvarbinary”、“image”、“clob”、“tinytext”、“longtext”、“text”、“longvarchar”、“char”、“nchar”、“nvarchar”、“nclob”。涵盖了所有常用的数据类型。

另外,还支持常用的数据库函数,如:ABS、ACOS、SUM等。在开发算法和底层模块时,只需要按照定义的标准SQL去编写操作语句。针对数据跨域问题,平台引入了临时库的机制(tmpDB),将需要的数据暂存到临时库的内存表(Memory Table)中,然后直接在临时库中运行标准SQL命令即可。由于数据都是保持在内存表中,运行效率高。

4.2 算法设置模块

平台中所有业务功能都是通过算法来实现,将算法单独提炼,构造算法设置模块,并将复杂算法分步骤拆分成多个简单节点,既提高算法的复用性,又能简化业务构建的复杂程度,提高数据平台的易用性。

水质评价算法中,设置了3个计算节点,包括:因子等级计算、超标倍数计算、营养状态等级计算。用户根据评价标准变化可以直接修改算法,提高易用性,同时平台提供了友好的节点改进界面,用户能在每个阶段也进行深入的改进。通过这些节点的修改,即可实现算法的修改,而不用直接修改复杂的原始程序,实现了用户对系统的可视化改进。各节点计算结果通过算法结果的前端展示界面查询,在此界面用户灵活自定义选择条件,包括:年份、月份、省份、数据类型、考核因子(一个或多个)等,点击“查询”,系统会自动调用上面算法,计算出评价结果,并在该页面展示出来。

5 平台应用总结

基于EAM的设计思想构建的水环境管理平台,构建了一种水环境数据库的规范化建设模式,有效解决了系统兼容、需求变化、数据安全、统计效率低下等问题。其优点如下。

(1)有利于系统之间的融合。对原有信息系统,平台能从底层数据库对这些系统进行集成,统一管理系统数据,使老旧系统正常运行。针对新的需求,利用该数据中心平台,均可以直接创建相应的算法,并创建相应的前端业务模块调用该算法,实现业务扩展。

(2)支撑业务变化和拓展。平台融合了业务数据,支持底层数据结构的扩展和变化,可以根据业务要求组装出符合需求的数据集合,从底层上支持业务的变更。同时,平台容易创建新的数据库对象,轻松应对新业务的开发。

(3)避免功能重复建设。数据管理平台是一个完整的系统,其逻辑功能模块提供了生长功能,用户实现了可视化的算法编辑,各种子系统的模块,只需在后台配置,不必重复开发。解决了当前的系统之间通信不完善,功能重复创建,造成资源浪费的问题,且由于构造的一致性,后期多系统融合较为简单。

(4)避免数据重复。数据中心的实现在不同功能与系统间的数据采取了统一管理,实现了数据统一。由平台托管数据,为多个业务系统提供数据服务,避免了不同系统中多重数据的问题。

(5)保障数据安全。数据管理平台直接建立了完备的数据管理与数据安全机制。具备存储操作备份功能,包括操作前的数据、操作后的数据、受影响的其他数据等,对误操作,如误删、误改,可以实施数据覆盖,保障数据安全。

但是,数据平台开发门槛高,系统开发难度大,受限于研究人员的时间和能力限制,目前水环境管理平台仍有较大改进空间,比如,平台并未集成数据的挖掘算法,应对海量数据的信息挖掘需求支撑不足,平台在数据校验的算法设计方

(下转第44页)

污许可证至少应按照不同行业、产业类型分为水污染物排放许可证、大气污染物排放许可证。许可证记载的包括管理节点的控制值和达标的判别标准、污染物排放值及执行的排放标准、监测方案、达标计划、信息记录要求、阶段报告要求及违法处罚等内容^[4]。

可见,排污许可证应当成为政府环境管理的执法文件、应当成为排污单位合法生产依法排放的执行文书,应当成为其他环境管理手段实施的基础和平台。其对污染源管理的引领作用降低了管理成本,提高了管理效率,决定了环境管理的有效性和环境质量改善的有效性。

3 环境管理制度体系的重构

排放源的管理是环境管理的根本,许可证制度担负着环境保护核心的管理职能,重塑排污许可证制度体制,理解排污许可证制度的重要地位,有效整合现行环境管理制度,实施一证式管理,代表着新型环境管理制度体系的形成。

随着生态文明体制改革的推进,环境保护行政主管部门与其他部门的管理界限和职责逐步清晰,未来环境管理逐步过渡并整合到3大体系。

(1)以排污许可证为核心的污染源管理体系。将污染源管理相关的环评审批、竣工验收、污染源监管、总量控制、排污收费、环境统计、信息公开等制度内容,逐步融合或替代,围绕许

可证制度构建新的源管理模式。

(2)基于监测、监察的环境安全和环境风险监督管理体系。独立于地方政府和环境管理部门,实施有效的监管和执法。

(3)基于环境承载力空间管控的环境质量控制顶层设计体系。区域承载力空间管控是环境质量改善的根本,通过多规合一,将环境管理要求融入到城市发展的顶层设计,将环境监管纳入到城市发展的过程中,建立空间环境管理体系代表着环境管理制度体系适应发展需求,融入“五位一体”发展总体布局。

其中,污染源管理体系是环境管理核心,环境监管体系是环境管理的保障,通过空间规划、多规合一实施的顶层设计体系是环境质量改善的出口。3大体系职责明确,边界清晰,突出体现了“谁污染,谁治理”、“全面监督”、“预防为主,防治结合”的管理理念,同时体现了将政治、经济、社会、文化融入到环境管理制度体系之中。

参考文献

(上接第30页)

面仍不足,应对多源数据的校验尚有较多改正的空间等。未来,将进一步优化相关功能,发挥平台作用,规范相关内容建设,推出数据库建设标准,促进数据库建设统一,促进我国水环境管理的优质数据库建立,为大数据分析奠定基础。

参考文献

- [1]习近平.关于《中共中央关于制定国民经济和社会发展第十三个五年规划的建议》的说明[N].人民日报,2015-11-04(02).
- [2]吴舜泽,吴悦颖,王东.将总量控制完善发展为生态文明建设基本制度[N].中国环境报,2013-11-12(02).
- [3]吴舜泽,洪亚雄.“十二五”环保规划布局[J].环境经济,2012(04):25-28.

- [1]郭逸飞,宋云,张彩丽,等.瑞典环境许可制度的特点分析及启示[J].环境保护,2015,43(6):62-64.
- [2]周宏春.新形势下我国环境管理与改革取向[J].中国环境管理,2012(11):8-13.
- [3]赵若楠,李艳萍,扈学文,等.排污许可证制度在环境管理制度体系的新定位[J].生态经济,2014,30(12):137-141.
- [4]宋国君,韩冬梅,王军霞.中国水排污许可证制度的定位及改革建议[J].环境科学研究,2012(9):1071-1076.
- [4]吴舜泽,徐敏,马乐宽,等.新形势下如何深化水污染防治工作与管理转型[J].环境保护,2014(11):35-38.
- [5]吴舜泽,王东,马乐宽,等.向水污染宣战的行动纲领——《水污染防治行动计划》解读[J].环境保护,2015(09):15-18.
- [6]魏斌,黄明祥.新形势下环境信息化发展展望[J].中国环境管理,2015(01):14-17.
- [7]吴舜泽,徐敏,马乐宽,等.重点流域“十三五”规划落实“水十条”的思路与重点[J].环境保护,2015(18):14-17.
- [8]徐洁.太湖流域水环境管理信息系统平台研究[D].南京:南京师范大学,2006.
- [9]林楠,冯玉杰,吴舜泽,等.我国跨区域水环境信息共享机制[J].哈尔滨工业大学学报,2012(12):41-46.
- [10]李玲璐.低成本流域空间数据库建库研究——以滇池流域为例[J].云南地理环境研究,2011(03):107-110.
- [11]李爱红,胡平,林宣雄.数据仓库在水环境质量评价系统中的应用前景[J].环境科学与技术,2004(03):52-54.