

基于元数据的油气信息管理系统^①

帅训波

(中国石油勘探开发研究院廊坊分院 地球物理与信息研究所, 廊坊 065007)

摘 要: 非结构化数据在信息总量中所占的比例远大于结构化数据, 是油气信息管理过程中的宝贵数据资产, 如何对这些非结构化数据有效管理和应用是新型油气信息管理系统建设的重要内容之一. 为了实现对油气信息资源中包括非结构化数据在内的综合数据高效管理与集成应用, 从应用元数据技术角度, 一方面对非结构化数据进行定义和详细描述, 另一方面以数据库元数据为核心对系统框架进行构建, 从而设计了一种新型的油气信息管理系统. 系统实现了油气信息综合数据的存储与管理, 而且可以对异构数据源进行有效集成应用, 有良好的灵活性和扩展性.

关键词: 元数据; 非结构化; 信息管理; 扩展

Oil and Natural Gas Information Management System Based on Metadata

SHUAI Xun-Bo

(Institute of Geophysics and Information, Langfang Branch of Research Institute of Petroleum Exploration and Development, Petro China, Langfang 065007, China)

Abstract: Unstructured data occupies the proportion of total information amount more than structured data, which is valuable data asset during management of oil and natural gas information. How to manage the unstructured data is an important content for developing new oil and natural gas information management system. In order to realize efficient management and integration application of unstructured oil and natural gas data, this paper puts forward a new oil and natural gas information management system from application of metadata technology view. In the new information management system, the unstructured data is defined and described by metadata, and the system framework is designed based on metadata of database. The new system has good flexibility and extensibility, which not only can efficient manage unstructured data of oil and natural gas, but also can realize integration application of heterogeneous data sources.

Key words: metadata; unstructured; information management; expand

1 引言

随着信息化和数字化进程在各个领域的持续推进, 以扫描图像、文档、演示文稿、语音和视频片段等为代表的非结构化数据在信息总量中所占的比例, 远大于以交易记录、流程控制和统计分析等为代表的结构化数据^[1,2]. 对油气信息管理过程中的大量非结构化数据进行有序存储和高效管理, 并尽可能地充分利用它们进行知识挖掘, 已成为全球各能源企业提高自

身核心竞争力的重要手段之一.

一般说来, 不易直接从非结构化数据中获取其蕴含的信息属性, 需要借助相应的专业软件工具才能将其打开浏览, 具有数据格式多样、数据量大、缺少较好的存储策略和难以检索等特点. 如果像结构化数据管理方法一样, 将非结构化数据直接存储于数据库中, 不仅大幅度增加数据库的容量外, 而且大大降低了数据库的运行和维护效率. 因此, 如何对信息管理过程

^① 基金项目: 中国石油天然气股份有限公司信息科研项目(2011-N62)

收稿时间: 2013-11-15; 收到修改稿时间: 2014-02-19

中非结构化数据进行有效管理和应用,成为新型信息管理系统建设的重要内容^[3-6]。

为了实现对油气信息资源中非结构化数据和结构化数据的高效综合管理与企业集成应用,本文从应用元数据技术角度,一方面对非结构化数据进行定义和详细描述,另一方面以数据库元数据为核心对系统框架进行构建,从而设计了一种新型的油气信息管理系统。系统实现了油气信息综合数据的存储与管理,而且可以对异构数据源有效集成应用,有良好的灵活性和扩展性。

2 元数据技术

元数据是描述数据的数据,对数据属性的信息进行描述,主要用于用户能够发现资源、识别资源、评价资源,能对相关的信息资源进行选择、定位和调用,并且可以追踪资源在使用过程中的变化,实现信息资源的整合、管理和存储^[4,6,7]。用户可以通过元数据查询和管理的方式,对数据库里的业务数据进行快速访问。常用的元数据操作有元数据选取与注册、元数据发布、元数据查询和元数据维护等。

为了更好地实现对油气非结构化信息的访问效率和安全管理,本文构建的元数据主要内容是对非结构化数据特征和属性进行详细描述,元数据基本内容抽象为<非结构数据,特征,特征值>的三元组形式,它可以对非结构数据的某一特征进行描述。依据数据库语言,元数据中的特征即是填入特征值数据的“字段”。如果对于一个非结构化数据,抽取到多个自身特征数据,那么,该非结构化数据可以由多个特征三元组组成的集合来进行详细描述。采取元数据集合来对非结构化数据进行详细描述,而没有选用<非结构数据,特征 1,特征 2,...,特征 N,特征值 1,特征值 2,...,特征值 N>的多元组形式,是因为元数据集合中一个三元组元素为一个特征描述,在对非结构化数据的特征查询和企业集成应用等方面具有更好的灵活性和扩展性。

从应用元数据的角度,采用管理结构化数据的方法,对非结构化数据的元数据进行有效管理,从而实现了非结构化数据进行整体管理和应用。例如技术报告文档是油气信息资源管理中的一种典型且重要的非结构化数据,这一类文档一般都包含文档名称、格式、完成单位、摘要等信息,因此,将名称、格式、完

成单位和摘要等作为描述技术文档的元数据,则描述一份技术报告的元数据集合为: {<技术报告,名称,“川中气田开发实施方案”>, <技术报告,格式,“Word 2007”>, <技术报告,完成单位,“中国石油勘探开发研究院廊坊分院,中国石油西南油气田分公司”>, ...},那么,在数据库表中建立相应的字段如表 1 所示:

表 1 技术报告的元数据

序号	名称	数据库字段
1	名称	Name
2	格式	Form
3	完成单位	Unit
4	摘要	Abstract

3 系统设计

3.1 数据管理

从用户应用功能角度,首先对油气信息数据划分为石油类和天然气类,然后对于每类数据又细分勘探类、开发类、采油(气)类和生产经营类等四大类,最后再按照油气生产管理的规范化数据控制流程,将每类数据再详细分为各种具体的结构化数据和非结构化数据,如何实现对这些非结构化数据和结构化数据的统一高效管理是本文新型油气信息管理系统设计的重要目标。在第 2 节部分,从应用元数据的角度,对非结构化数据的特征和属性进行了详细描述。选用关系数据库模式对结构化数据和非结构化数据的元数据进行管理,但对非结构化数据的物理存储管理上,采用了文件管理方式,将其存储在磁盘上,文件读取的入口地址作为元数据的一个特征值存储在数据库中。为了使系统具有良好的灵活性和扩展性能,系统中对结构化数据也采用元数据技术进行管理,建立了结构化数据的数据字典。

为了实现对非结构化数据的动态快捷操作,确保数据库中元数据信息实时地反应出非结构化数据的动态变化情况,在数据库设计环节,对非结构化数据文件设计一个服务元数据关系表,它主要负责对非结构化数据的元数据检索、定位和实时管理,对非结构化数据文件的动态更新操作后,在服务元数据表中对操作结果实时记录。本文对数据管理的模型示意如图 1 所示:

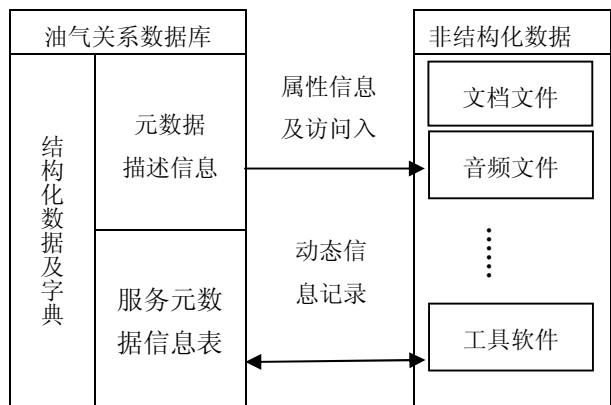


图 1 数据管理模型示意

3.2 系统架构

考虑到系统自身应具有良好的灵活性和易维护性，兼顾将来与其它应用系统进行企业集成应用的可行性。系统架构的设计基本思想是，应用程序通过对各类数据的元数据访问来获取数据，建立面向元数据应用的各类功能方法。当系统管理的数据对象发生改变时，只需修改相应的元数据集合，改变了传统信息

管理系统中不仅要修改底层数据库，而且要改动对应的功能程序的做法。

以元数据管理和应用为核心，本文设计了三层系统架构，即数据管理层，元数据管理层和数据应用层，结构如图 2 所示：

数据管理层存储各种信息系统数据，按照一定格式将油气生产过程中产生的非结构化数据文件、结构化数据库文件和元数据信息等存储管理。从而构成系统的底层数据库。

元数据管理层对从数据管理层抽取的所有元数据进行注册、发布、查询和应用维护。用户通过能元数据查询和管理的方式访问数据库的油气生产业务数据。

数据应用层主要完成信息数据的采集入库，通过服务元数据应用接口对数据的动态更新进行发布，按照用户对数据的应用权限规范对数据综合应用，它是访问信息数据的用户界面和接口。

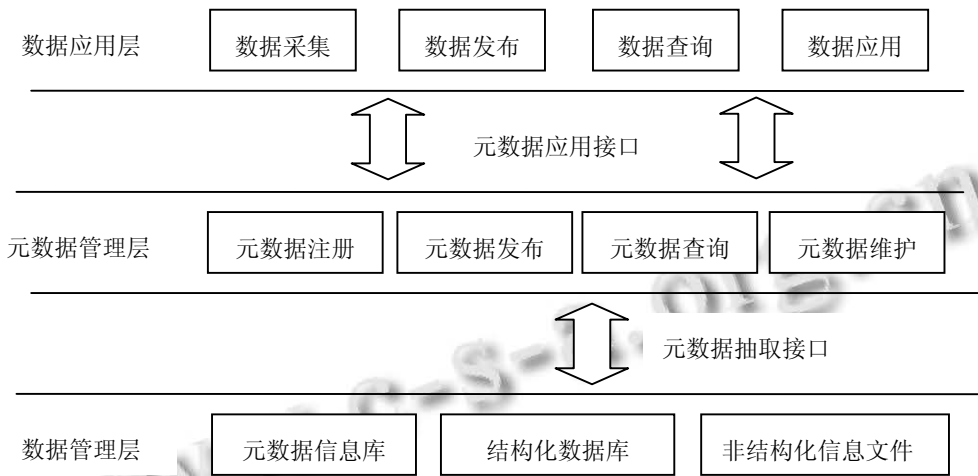


图 2 系统架构

3.3 应用功能

本文所研发油气信息管理系统在应用功能设计方面，以油气区块的综合数据管理与应用为基本依据，按照数据管理流程，系统的用户应用功能分为了数据采集与查询、动态分析和决策支持等 3 个主要子系统，应用功能模块结构如图 3 所示。

数据采集和查询子系统包括计划数管理、生产周报管理、动态数据管理、静态数据管理、HSE 数据管理和过程文档管理等 6 个主要功能模块，主要是完成

油气信息综合数据的采集和查询功能。所包括的结构化数据主要是油气区块和单井数据信息为主，非结构化油气信息数据综合应用主要通过 HSE 数据管理和过程文档管理两个功能模块与其它功能模块交互通信实现。

动态分析子系统主要包括开发动态预测、开发指标概算、开发动态分析和开发效果评价等 4 个功能模块。将常用的各种油气开发动态预测方法进行封装，通过调用数据采集和查询子系统查询功能，获取

相应的动态分析数据参数,对油气区块生产动态进行分析,将动态分析结果以图形曲线的形式展现。

决策支持子系统主要包括运营类数据分析、计划类数据分析、收益率分析和 HSE 数据分析等 4 个主要功能模块,以系统内的年度综合管理数据为基础,进行评价分析,为油气区块的生产决策提供信息支持。其中,运营类数据分析主要包括油气储量价值分析、油气成本分析、产销量分析和钻井成本分析等。在本系统中,数据采集和查询子系统实现了非结构化

数据的管理。由于在动态分析子系统和决策支持子系统中通过元数据应用接口,能获取到传统油气信息管理系统中没有包含的非结构化数据信息,从而增强了分析方法所需参数数据的客观性,因此,该系统所提拱的动态预测效果更能进一步反应出油气区块的实际生产变化情况,所提供的决策支持信息具有更好的依据性和科学性,推广应用效果也充分证实了该系统的优良特性。

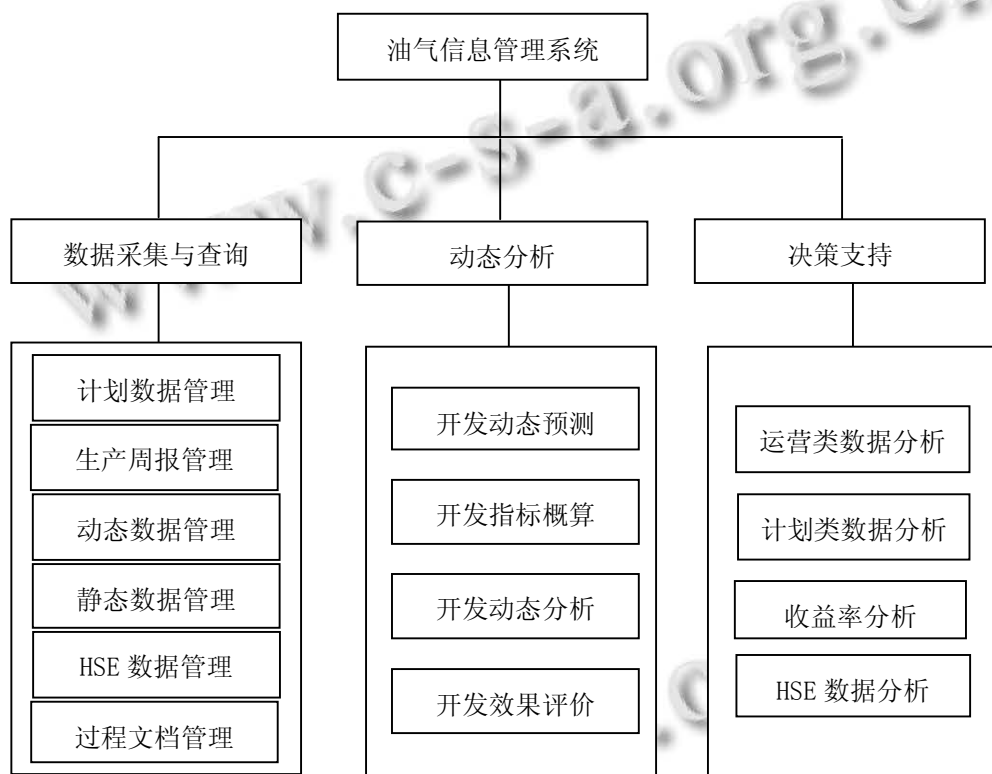


图 3 系统应用功能模块结构

4 系统实现的关键技术

4.1 Web Service 数据集成

第 2 节论述了应用元数据技术对非结构化数据进行特征属性定义和抽取。XML 作为一种可扩展标记语言,具有结构规范、语法简明和应用灵活等优点^[8,9],因此,元数据的表示方式采用 XML 格式,元数据的多元组可以由 XML 文档中的 XSD 文件来定义实现。

Web Service 是基于网络的、分布式的模块化组件,它采用标准化的 XML 消息传递机制实现与其兼容的组件进行互操作^[8,10],因此,系统中数据管理层与元数据管理层、数据应用层与元数据管理层等之间的通信

实现选用 Web Service 技术。在系统的每个应用层,应用 Web Service 实现最基本的数据访问功能,将元数据及其功能封装起来,实现元数据的应用接口,供外部调用。Web Service 的接口规范包括参数数据格式、返回结果格式和返回结果的元数据描述等。

从应用发展的角度分析,系统需要与其它系统集成时,由于采用 Web Service 技术开发,具有规范的数据通信接口,能够较容易地实现异构数据源的集成应用。因此,从技术应用层面来讲,系统自身具有良好的开放性和易扩展性。

4.2 数据访问控制

为了加强系统的数据应用安全,从应用元数据技术基础之上^[11,12],引入基于角色的访问控制技术,将二者应用结合起来,实现对数据的访问权限控制.首先根据系统的元数据表,划分为各种的数据应用权限;然后根据油气信息管理业务需求,将每种权限确定为对应的“岗位”所拥有;最后再按照业务管理中的“岗位”进行分类,每类岗位所拥有的元数据操作功能的权限集合确定为系统中一个角色.在系统应用功能中,将给定的用户账号赋予一个或者多个角色,该用户账号便拥有角色的所有权限集合,从而拥有了该权限集合中元数据所对应的生产综合数据访问控制权限.

系统的实现采用上述基于元数据和角色管理相结合的数据访问控制方法,通过元数据的管理,将底层生产数据与使用权限相分离.当元数据被修改时,系统中数据管理权限也被相应变化.通过对角色的管理,将用户和访问权限相分离.当用户所拥有的角色发生改变时,其拥有的数据访问控制权限也就发生了相应的改变.由此可见,系统中所采用的基于元数据的角色访问控制技术既确保了系统具有良好的安全性,又具有很好的灵活性.

5 结语

本文从应用元数据技术角度,一方面对非结构化数据特征和属性进行详细描述,另一方面以数据库元数据为核心对系统框架进行构建,系统中应用功能程序通过对元数据的访问来获取数据,从而建立了面向元数据应用的功能方法.以油气区块的综合数据管理与应用为基本依据,按照数据管理流程,将系统的用户应用功能分为了数据采集与查询、动态分析和决策支持等 3 个主要子系统,并采用 Web Service 技术实现.将基于角色的访问控制与元数据管理相结合应

用,加强了系统中的数据访问安全.推广应用效果表明,系统对油气信息资源的综合应用管理有着较好的效果,且具有良好的扩展性.

参考文献

- 1 詹炜,万琼.基于元数据的企业信息数据仓库系统构建.长江大学学报(自然科学版),2010,17(9):569-570.
- 2 帅训波,买炜,魏东,邵艳伟.天然气综合管理信息系统设计与开发.菏泽学院学报,2011,33(2):32-34.
- 3 步红.基于元数据的指挥信息管理及共享技术研究.网络安全技术与应用,2011,11(9):11-13.
- 4 文必龙,李智新,王英艳.基于元数据的企业搜索引擎研究.郑州轻工业学院学报(自然科学版),2008,23(6):4-6.
- 5 刘实,郭俊芳,刘杉,沙仁高娃,田鹿,郝晓琴,岳鹏飞.基于元数据的教育资源领域本体的研究与构建.中山大学学报(自然科学版),2009,48(S):271-276.
- 6 李颖,高健翔,贾宇田.元数据技术在油气资源数据共享服务中的应用.辽宁工程技术大学学报(自然科学版),2012,31(3):383-386.
- 7 袁满,陈星童.基于元数据驱动的异构数据模型映射算法.东北林业大学学报,2011,39(12):128-131.
- 8 张宇,蒋东兴,刘启新.基于元数据的异构数据集整合方案.清华大学学报(自然科学版),2009,49(7):1037-1040.
- 9 李焕勤.用 XML 在 Word 文档表格中转换非结构化数据.计算机与应用化学,2010,27(9):1297-1299.
- 10 帅训波,关新,卢明,李家庆,吴兵,周相广.基于 WEB 的中石油项目管理系统研发.天然气技术,2008,2(1):84-88.
- 11 陈小青,马亚平,段永强.基于主题映射元数据的数据库访问和集成系统.系统仿真学报,2009,21(14):4268-4273.
- 12 王晓超,赵卫东,左青香.基于元数据和角色控制的用户权限管理.计算机技术与发展,2012,22(3):233-236.