

大数据时代地质灾害数据管理及应用模式探讨

马 敏*

武汉地质调查中心 湖北 武汉 430205

摘 要：地质数据是一种多源异构数据，具有数据量庞大、数据结构多样等特点，属于大数据范畴。在地质数据中，非结构化数据占有很大的比例，关系模型难以有效地对它们进行管理，这种弱点，在大数据时代将会被进一步放大。针对大数据时代地质数据的管理模式，本文从原始数据索引管理、泛结构化数据异构数据库管理、中间件集成等方面进行了探讨。

关键词：大数据；泛结构化地质数据；数据管理

DOI：https://doi.org/10.37155/2717-5189-0311-44

1 地质大数据管理及应用模式

1.1 原始数据索引管理

通过地质调查、监测等手段获取的各类原始数据，一般要经过大量的结构化、标准化、归一化等的处理，才能建立关系型地质数据库。在这一过程中，原始数据实际上经过了大量不同程度的分解、提取、转换、量化等操作，造成关系型数据库中的数据与原始数据之间可能产生不同程度的差异，这种差异会对地质分析的结果造成不同程度的影响。除规范工作流程以外，原始数据索引管理在一定程度上可以减弱这种差异。原始数据文件索引管理是指在建立传统的关系型地质数据库时，也同时建立原始数据的索引管理。简单方法是：①把原始数据分门别类地存储在不同的目录中，存储的原始数据没有经过任何改动；②建立文件检索数据表对原始数据进行管理。文件检索数据表可以快速定位并调用所有的原始数据文件^[1]。

1.2 泛结构化地质数据管理

地质数据是一种泛结构化大数据，除了通过传统关系型数据库进行管理的结构化数据外，还有两类数据是地质数据中常见的类型：一种是文档类数据，如大量的地质报告^[2]。另一种是各种地质图件。由于关系模型对非结构化数据内部包含的丰富信息无法有效控制，因而针对非结构化地质数据的NoSQL技术近几年来受到广泛地关注。结合NoSQL技术的泛结构化地质数据管理途径如下（图1）。

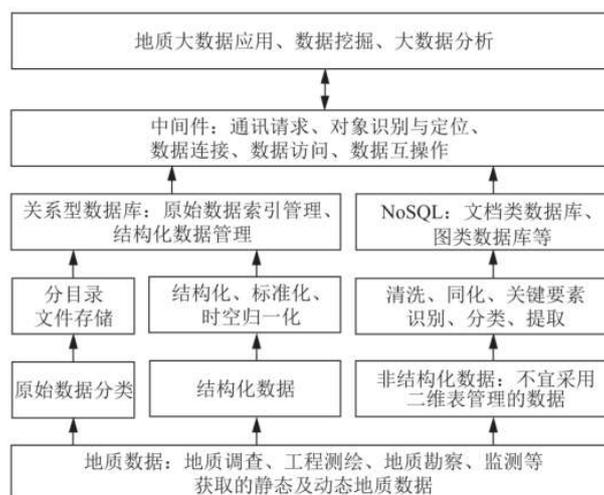


图1 泛结构化地质大数据管理模式

(1) 对获取的原始地质数据进行分类，如把原始数据分为钻探原始数据、测绘原始数据、试验原始数据等，建立

*通讯作者：马敏，女，汉族，湖北武汉，硕士，中级员工，武汉地质调查中心，研究方向：地质信息技术。

对应的目录，把获取的原始数据分门别类地存放。然后建立原始数据文件的索引记录表，记录对应文件的文件名称、工程名称、勘察阶段、工作单位、存放位置、存放时间等。文件索引数据表通过关系型数据库进行管理。（2）从地质数据中分解出传统的多种字段，对地质数据进行二维数据表结构化处理，字段标准化处理，以及统一时空坐标处理等，建立传统的关系型地质数据库，如SQL Server地质数据库等。这种管理方式是当前地质数据管理的主流方式。

（3）从地质数据中获取地质报告、地质图件等非结构化地质数据，使用NoSQL来进行针对性管理。

相对于关系型数据库，NoSQL是一种新的数据管理模式。比如我们可以采用文档型No-SQLMongoDb来管理地质报告类数据等。NoSQL一般适用于比较简单的数据模型，需要很高的灵活性，对数据库性能要求较高，不需要高度的数据一致性。对于给定键值（key），比较容易映射复杂值（Value）的环境。

1.3 中间件集成

地质泛结构化数据管理需要关系型数据库和多种NoSQL支持，它们属于异构数据库，其数据组织、数据访问与数据提取的方法是不同的。即在地质数据管理及应用过程中，需要多种系统的共同支持，多种异构系统的集成通过中间件来实现。在地质大数据处理的分布式环境中，中间件可以连接多个应用程序或系统，通过网络通讯实现不同系统之间的信息传递、数据交换和数据互操作。中间件需要屏蔽分布环境中异构的系统和网络协议，提供分布环境下的通讯服务。在地质大数据管理中，中间件在异构数据库之间、泛结构化数据与不同应用系统之间提供数据的通讯请求、对象识别与定位、数据连接、应用申请及申请结果返回等功能，实现异构数据与多种应用系统之间数据的互操作（图2）。

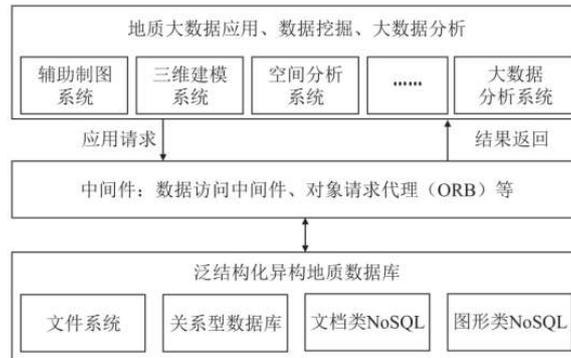


图2 地质大数据管理及应用集成方案

泛结构化地质数据库中的文件系统、关系型数据库和NoSQL数据库都可以有多个，如多个工程，每个工程的原始数据都可以是一个文件系统。关系型数据库也可以分为勘察数据库、监测数据库等。NoSQL也可以依据不同类型、不同工程分为多个。地质数据应用分析、数据挖掘、大数据分析也可以分为多个应用系统，如辅助制图系统、三维建模系统等，不同应用系统需要不同数据集合的支持。应用系统首先依据地质专业需求向中间件提出数据需求申请，中间件访问泛结构化地质数据库，从其中检索并提取所需的数据或数据集合，再以应用请求需要的格式返回给提出申请的应用系统，完成对数据需求的操作^[3]。

2 几点讨论

地质大数据的管理采用文件系统、关系型数据库及NOSQL数据库 3 种并行的管理方式，可以有效管理泛结构化地质大数据。但是这样必然造成数据管理的异构性，即地质数据的管理是通过多种不同结构的数据库来进行管理的，这将给数据的集成及数据的互操作带来更大的压力。这种压力可以通过中间件技术克服。NoSQL在地质数据管理中可以起到如下几个作用：（1）管理地质数据中的非结构化数据，可以对包含在非结构化数据中的数据进行检索、提取等操作；（2）通过快速获取包含在非结构化数据中的数据，弥补地质大数据分析中只依靠关系型数据库提供数据的局限性；（3）由于NoSQL可以对包含在非结构化数据中的数据进行操作，因而可以针对包含在非结构化数据中的数据和信息进行多种数据再分析工作。

3 结束语

大数据理论与技术的发展，对传统的地质数据管理与分析造成一定的影响。通过原始数据索引管理、泛结构化数

据异构数据库管理及中间件集成技术,探讨了大数据时代地质大数据管理及应用的模式。认为采用文件系统、关系型数据库和NoSQL3种方式可以全面有效地管理泛结构化地质大数据,采用中间件技术可以很好地实现地质数据管理与分析过程中多种异构系统间的数据访问与数据互操作,这种管理模式是一种良好的地质大数据管理与应用模式。

参考文献:

- [1]吴润泽,程温鸣,刘军旗,等.三峡库区地质防治信息系统及预警指挥系统数据管理模式探讨[J].中国地质与防治学报,2018,29(5):102-107
- [2]吴广君,王树鹏,陈明,等.海量结构化数据存储检索系统[J].计算机研究与发展,2012,49(增刊1):1-5
- [3]罗颖.针对面向多源异构数据的数据集成中间件的设计与开发[J].网络安全技术与应用,2019(6):55-57