 DOI: https://doi.org/10.12345/smg.v7i3.26826

Research on the Organization and Management of Geological Big Data for Urban Underground Space

Renwu Fan[1] Baoyi Zhang[2] Binghai Wang[1] Qingzhong Zhan[2] Min Lu[1]

1. Institute of Geological and Geographic Information of Hunan Province, Changsha, Hunan, 410000, China
2. School of Geosciences & Info-Physics, Central South University, Changsha, Hunan, 410083, China

Abstract

In the development of urban underground space, the management and application of geological big data have profound implications. Confronted with challenges such as voluminous data, suboptimal utilization efficiency, an absence of a comprehensive governance framework, and difficulties in information extraction, this paper constructs a coupled data model and a distributed storage technology system for geological big data. The geological big data of urban underground space are stratified, graded, and classified based on professional and disciplinary fields, data access frequency, and data structure characteristics. Furthermore, a distributed storage optimization approach incorporating a hierarchical storage strategy is designed to realize the integrated management of multi-source and heterogeneous geological data. By leveraging relevant technologies, the data search efficiencies are improved, which provides new perspectives for the efficient management and utilization of geological big data in urban underground space and facilitating the development of urban underground space.

Keywords

Geological Big Data; Organization and Management; Multi-source and Heterogeneous; Distributed Storage; Performance Optimization; HDFS; Urban Underground Space

面向城市地下空间的地质大数据组织管理研究

樊任武[1] 张宝一[2] 王宾海[1] 詹庆忠[2] 陆敏[1]

1. 湖南省地质地理信息所（湖南省地质大数据中心），中国 · 湖南 长沙 410000
2. 中南大学地球科学与信息物理学院，中国 · 湖南 长沙 410000

摘 要

城市地下空间开发中，地质大数据的管理与应用影响深远。针对其数据量大、利用低效、治理体系缺失及挖掘困难等问题，本文构建地质大数据耦合数据模型与分布式存储技术体系，按照专业与学科领域、数据访问频率、数据结构特性对城市地下空间地质大数据进行分层、分级、分类，并设计融合分级存储策略的分布式存储优化方法，实现多源异构地质资料的一体化管理，借助相关技术提升数据检索效率，为城市地下空间地质大数据的高效管理与利用提供新思路，助力城市地下空间开发。

关键词

地质大数据；组织管理；多源异构；分布式存储；性能优化；HDFS；城市地下空间

1 引言

城市化进程的加速使得城市空间日益紧张，地下空间作为重要的城市资源，其开发利用正成为全球范围内的一个重要发展趋势[1]。城市地下空间的开发利用潜力巨大，并且涉及到多种资源和功能的整合。地铁及地下道路网络可有效缓解地面交通拥堵问题，地下停车场的广泛布局能够减少路面停车对公共空间的占用，进一步优化城市交通环境[2]；而地下仓储物流中心的建立则有助于改善城市的货物配送体系，减少地面运输压力[3]；利用地下空间进行商业开发（如购物中心、娱乐设施等）不仅能够提升土地利用效率，还能为市民提供更加便捷的生活服务[4]。科学合理地开发利用城市地下空间不仅可以解决城市发展过程中面临的诸多挑战，还能够促进城市的可持续发展，增强城市的综合承载能力和竞争力。

随着城市地质探测技术的进步，城市地下空间的地质大数据资源呈现指数级的增长。地下钻探技术、浅层地震、

【基金项目】湖南省自然资源厅科技计划项目（项目编号：湘自资科 20230123XX）；湖南省地质院重点科研项目（项目编号：HNGSTP202301）。

【作者简介】樊任武（1976–），男，本科，从事湖南省地质大数据建设，建立湖南省地质云数据计算、管理和公益查询服务体系等研究。

高密度电法、探地雷达、微动等地球物理学方法的广泛应用，使得对地层结构、岩土特性、地质构造等地质信息的获取更为精准和全面[5,6]。然而，海量地质数据的涌现也带来了新的难题，如何有效整合、分析和利用这些数据，成为当前亟待解决的问题。若不能妥善处理这些数据，不仅会造成资源浪费，更会阻碍地下空间开发项目的高效推进。例如，在规划地下工程时，因无法从繁杂数据中精准提取关键地质信息，可能导致工程选址失误、施工风险预估不足等问题[7]。

因此，深入研究地质大数据的组织管理与应用，对于充分挖掘城市地下空间开发潜力、提升开发效率与安全性、实现城市可持续发展具有至关重要的意义。本文针对城市地下空间的地质大数据资源量大、利用效率不高、缺乏有效的大数据治理体系和信息挖掘困难的现状，构建了耦合数据模型及分布式存储技术体系，实现了多源异构数据一体化管理。提出融合分级存储策略的分布式存储优化方法，通过空间索引与高效检索技术，提升多访问模式下的异构数据存储与访问效率，为城市地下空间地质大数据的高效治理提供技术路径。

2 数据组织与存储管理

针对城市地下空间地质大数据，按照专业进行数据分层、按照数据结构进行数据分类、按照访问频率进行数据分级。本文主要对水文、工程、环境三个专业领域进行数据分层；进一步按照数据访问频率进行分级存储；根据数据的结构特性确定数据库与存储管理策略。

2.1 数据的分层分类分级组织

城市地下空间地质大数据按照分层、分类、分级的形式进行组织管理，如图 1 所示。本文主要对水工环的地质数据进行分级管理，根据数据的使用频率、数据类型以及存取需求，确定分级分类存储策略，以平衡存储成本和数据访问效率。分级存储主要表现为不同类型的数据根据访问频率划分为“热数据”（经常访问）、“温数据”（适度访问）和“冷数据”（很少访问）。分类存储表现为根据数据的格式和结构特性，划分为结构化数据（如数据库中的表格数据）、半结构化数据（如 JSON、XML 等）以及非结构化数据（如图像、视频、文本等）。

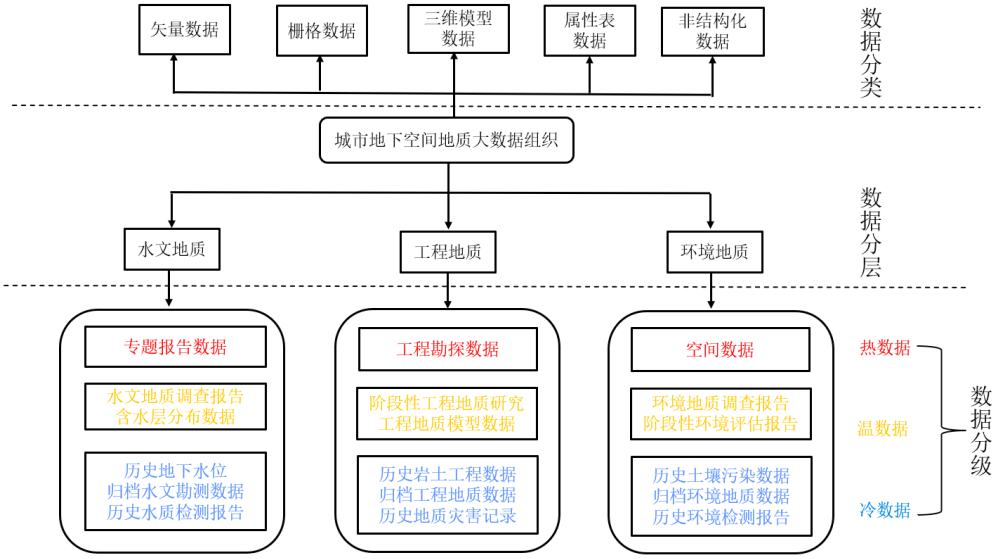

图 1 城市地下空间地质大数据组织形式

2.2 分级数据的存储与迁移备份

针对不同数据的特点设计分级存储架构。热数据使用快速的分布式存储系统，如 ElasticSearch，确保实时数据处理和查询的高效性，包括地质图、地形图等空间数据，柱状图、剖面图等勘探数据，遥感影像数据，以及成果报告、表格、图件和专题报告等成果数据。温数据选择关系型数据库（如 MySQL 或 PostgreSQL）或中等性能的 NoSQL 数据库（如 Cassandra、MongoDB）来管理和存储定期访问的数据，包括地质灾害监测等动态实时数据、分析测试数据、地质资料数字化的数据。冷数据利用分布式文件系统（如 HDFS）进行存储，以保证对大规模文件的高效管理和快速访问，包括地质调查报告、地质文献等文本数据，档案资料，野外调查、工程揭露等形成的各种原始过程资料。

针对数据迁移与备份，遵循定期原则将热数据进行快照备份，并将其迁移到冷数据存储中，形成冷备份，以提升系统的存储效率、防止数据丢失。在数据备份时，采用基于云的存储备份服务，确保数据的安全性和持久性。

2.3 分类数据的管理

从信息技术处理的角度，可以将水工环的结构化、半结构化、非结构化地质数据，重新划分为关系型数据、非结构化数据、空间数据和知识图谱 / 三维模型四类。同时构建数据中控层，实现统一监控、分类和处理流程的标准化操作；并根据不同数据类型的特点选择合适的数据库存储技术，在性能和成本上达到最佳平衡。

2.4 数据的分布式存储

对于文本 / 视频 / 影像 / 监测等非结构化数据，需支持 HDFS 作为分布式文件系统，以及 MapReduce、Hive 等工具进行数据分析和处理。本文构建了主从架构的分布式文件系统进行城市地下空间数据文件的存储和索引。对分层分级的水工环数据进行数据分片并创建空间索引，便于快速检索数据；随后为数据分配存储节点，对数据进行冗余存储。

3 数据访问与索引模式优化

3.1 数据访问模式与性能

在地质大数据并行处理场景中，当分布式节点存储的异构数据文件需要跨节点访问时，频繁的远程数据请求及任务间数据交互容易显著增加网络传输开销，影响系统整体性能。本文针对地下空间分析任务的空间拓扑特性和并行计算需求，基于数据关联特征与多尺度需求两个维度构建存储优化模型，以提升多模态访问场景下的分布式数据读取效率。

（1）数据关联特征分析。主要体现在要素相关性和空间相关性两个方面。一是多源地质探测数据的融合分析要求对同构地质要素建立协同访问机制；二是空间邻域内的数据文件在并行任务执行过程中呈现出较高的共现访问概率 [8]。

（2）多尺度访问需求解析。不同地下空间分析任务在空间尺度、数据密度和处理规模上存在显著差异：大尺度分析往往需要多计算节点协同访问 HDFS 集群，而轻量级查询则可依托单节点完成，这对数据分块策略提出了差异化要求，具体表现为分析范围的空间跨度、数据粒度的分辨率等级等参数存在数量级差异 [8]。

基于上述数据访问特征，本文引入以下三种访问模式：

（1）大尺度协同访问。基于空间拓扑关联的节点亲和存储机制，采用元数据标注实现关联文件在计算节点的共位存储，通过增强数据局部性降低跨节点通信负载。

（2）标准分块访问。采用全局 - 本地二级索引架构，当数据关联性较弱时，直接基于 HDFS 原生分块策略实施分布式访问，适用于非耦合型分析场景。

（3）小粒度融合访问。通过动态文件合并与空间聚类写入策略，构建基于空间密度的复合数据块，结合局部索引优化减少随机磁盘 I/O，提升细粒度查询效率。

应用上述三种数据访问模式，数据的访问性能得到了显著提升，如图 2 所示：

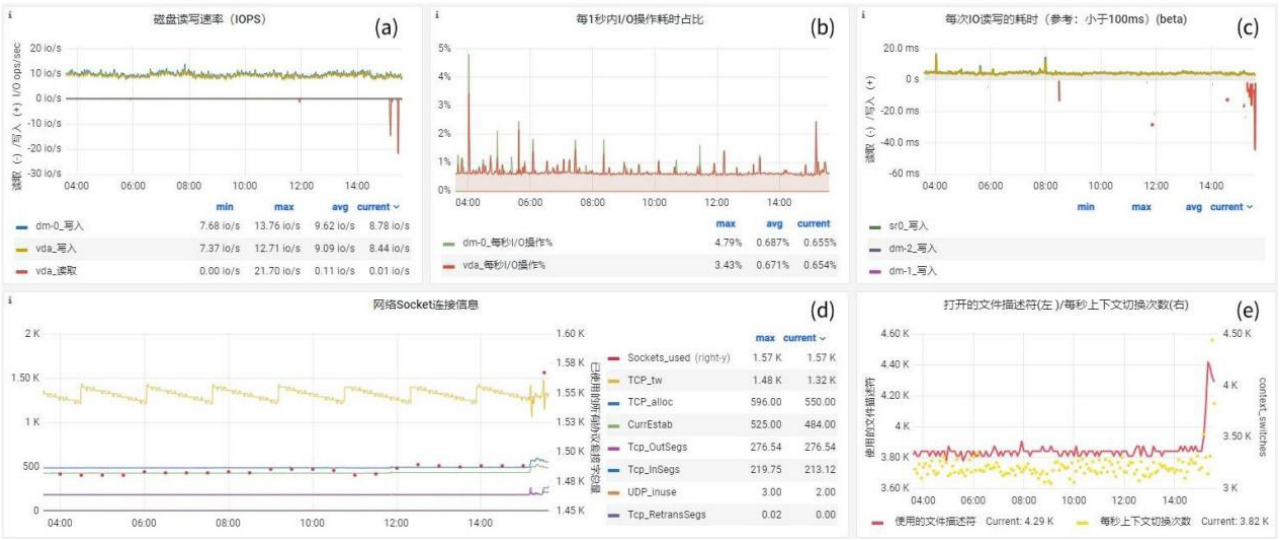

图 2 数据访问性能：（a）磁盘读写速率，（b）每秒内 I/O 操作耗时占比，（c）每次 IO 读写耗时，（d）网络 Socket 连接信息，（e）打开文件描述符 / 每秒上下文切换次数

3.2 数据索引模式与性能

针对大规模城市地下空间地质大数据，建立高效存储与检索的空间索引体系。根据空间范围将原始数据划分为 sblock，保存为一个独立的 HDFS 文件，然后 HDFS 自动将该文件切分为固定大小的 block，分散存储到集群节点，最后将 sblock 的空间范围与 HDFS 文件路径、block 列表绑定，写入全局索引。以保证相邻空间的数据尽量存储在同一节点，减少跨节点访问。对于需要建立更细粒度空间单元的数据，对 sblock 进行进一步细分，通过 R* 树索引结构生成 esblock，并为其分配唯一 ID，通过元数据将 esblock 的 ID、空间范围、物理存储路径关联起来，实现全局索引—HDFS 文件—sblock—esblock 的空间索引构建。

通过多层索引减少全量扫描，降低 I/O 开销，利用 HDFS 分布式特性支持 PB 级数据，运用 esblock 适配高精度的空间区域划分，提升了大规模数据的存储和检索效率，执行索引时系统的性能如图 3 所示。

4 结论与展望

在城市地下空间开发利用需求不断增长的背景下，地质大数据的科学管理与高效应用成为关键。本文聚焦于城市地下空间地质大数据资源现存的量大难管、利用低效、治理体

系缺失及信息挖掘困境等问题，展开了深入且系统的研究。通过构建城市地下空间地质大数据耦合数据模型及分布式存储技术体系，实现了各类地质数据的一体化整合，解决了多源异构数据空间一致、时间对齐、检索时间长、综合利用困难等问题，显著提升了数据管理的系统性与规范性。按照专业与学科领域、数据访问频率、数据结构特性进行城市地下空间地质大数据的分层、分级、分类，并进一步确定了数据的存储方案。进一步地，针对分布式存储系统的访问与检索瓶颈，确定了优化方法，融合分级存储策略，兼顾多访问模式与异构数据存储需求，有效增强了系统的访问灵活性与检索效率。借助分布式存储、空间索引与高效检索技术的协同作用，大幅提升了城市地下空间地质大数据的处理能力。

未来，还将引入 AI、数据挖掘等前沿技术，深化跨模态数据融合，提升城市地下空间地质大数据的自动化处理与智能分析水平；同时优化分布式存储系统的容错性与扩展性，以应对不断增长的数据量与复杂多变的应用场景。此外，还将探索与城市智慧化建设的融合路径，助力城市地下空间的可持续、智慧化开发利用。

图 3 执行索引性能：（a）资源明细，（b）各分区可用空间，（c）每小时流量，（d）CPU 使用率，（e）内存信息，（f）每秒网络带宽使用

参考文献

[1] 彭芳乐, 乔永康, 董蕴豪, 等. 新发展阶段城市地下空间开发利用发展战略研究[J]. 中国工程科学, 2024, 26(3): 176-185.

[2] 王涛. 地下交通系统构建与城市地下空间开发互动研究——以南通市为例[C]//中国城市规划学会城市交通规划学术委员会.协同发展与交通实践. 江苏: 2015年中国城市交通规划年会暨第28次学术研讨会论文集.江苏省城市规划设计研究院(江苏省城市交通规划研究中心), 2015: 97.

[3] 陈美好. 地下物流仓储中心定位技术研究[D]. 北京邮电大学, 2021.

[4] 李卓聪. 自然资源部发布《指导意见》推进城市地下空间有序开发利用[J]. 资源与人居环境, 2024(10): 10.

[5] 侯文仁. 废弃钻井液处理方法发展研究[J]. 测绘与地质, 2024, 6(1): 34-36.

[6] 崔静林, 袁珍, 董贵斌. “两宽一高”地震勘探技术在老气田中的应用——以哈萨克斯坦K气田为例[J]. 测绘与地质, 2023, 5(6): 105-108.

[7] 杨友刚, 雷世兵. 青藏高原易贡藏布流域水电开发重大工程地质问题及对策[J]. 工程地质学报, 2024, 32(5): 1759-1770.

[8] 刘志鹏. 城市地下空间大数据分布式存储与调度方法研究[D]. 中国地质大学, 2022.