文章编号:1671-4598(2023)06-0156-06 DOI:10.16526/j.cnki.11—4762/tp.2023.06.024 中图分类号:TP391 文献标识码:A

基于大数据应用的地质灾害数据存储策略

石晓栊[1],赵统永[1],王耀忠[1],彭 君[2]
(1.国家管网集团川气东送天然气管道有限公司,武汉 430079;
2.吉林大学 计算机科学与技术学院,长春 130012)

摘要:针对人工智能算法和大数据技术在地质灾害监测和预警上的应用需求,基于分布式文件系统(HDFS)和列式存储非关系型数据库(HBase)提出了地质灾害相关数据的存储策略;分析了地质灾害监控系统、地质灾害预测预报系统所需使用数据的数据种类、数据格式、数据容量、数据频率及数据增长速度等信息;从数据粒度大小的角度来对数据进行分类和组织,对不同粒度的数据设计了不同的存储模式,以实现高效的存取效率;根据数据的应用特性对数据进行类别划分,为不同类型的数据提供不同的存储结构和访问接口,以获得最优的数据访问性能。

关键词:监测预警;HDFS;HBase;分布式数据库;大数据应用;地质灾害

Geological Disaster Data Storage Strategy Based on Big Data Application

SHI Xiaolong[1], ZHAO Tongyong[1], WANG Yaozhong[1], PENG Jun[2]
(1. PipeChina Group Sichuan East Natural Gas Transmission Company, Wuhan 430079, China;
2. College of Computer Science and Technology, Jilin University, Changchun 130012, China)

Abstract: Aimed at the application of artificial intelligence and big data technology in geological hazard monitoring and forecast, a geological hazard data storage strategy based on distributed file system (HDFS) and column storage non-relational database (HBase) is proposed. The geological hazard prediction system and the geological hazard monitoring system information of data type, data format, data capacity, data frequency and data growth rate are analyzed. The data are classified and organized from the perspective of data granularity, and different storage modes are designed for different granularity data to achieve the high access efficiency. The data are classified by the application characteristics of data, and different storage structures and access interfaces are provided for different types of data to obtain optimal data access performance.

Keywords: monitoring and forecast; HDFS; HBase; distributed database; big data application; geological hazards

0 引言

随着人工智能和大数据技术的不断提高和成熟,在地质灾害监控、分析评估和预报预警等领域也获得了较为广泛的应用。然而数据作为人工智能和大数据技术应用的基础,数据的使用在数量上、质量上和频率上都被提出了很高的要求,传统的关系型数据库已无法适应数据急剧增长和应用模式变化带来的挑战[1-2]。

基于大数据应用的分布式非关系型数据库在此环境下应运而生。Hadoop生态圈的不断完善及HBase的发展和推广[3-5],为人工智能和大数据技术应用的拓展和深入提供了便利条件。蒋叶林、邹喆等在使用HBase进行时空大数据的处理方面进行了探索[6-7],王普刚、温静则将HBase应用到了实际的工程项目中[8-9],这些研究和探索证明了HBase良好的性能。

在对地质灾害相关的AI应用和数据服务方面,涂美义等基于SOA构建了省级地质灾害应急服务数据服务体系[13],张茂省等提出了基于人工智能的地质灾害防控体系[14],王晓刚等采用HDFS和HBase完成了洪灾数据的存储和管理[15],王键键等尝试了将大数据技术应用于城市洪涝灾害分析预警[16]。这些研究都涉及了人工智能和大数据技术对地质灾害相关数据的使用,但是没有形成一个较为完整、统一的数据支撑体系。

1 地质灾害数据类别分析

地质灾害监测数据包括:地质灾害点数据、巡查巡测数据、动态监测数据、预警预报数据、报告类数据、地质/地理数据、气象数据、管道本体数据、多媒体数据等类别。不同类别的数据,具有的格式、属性、大小、数据量、更新频率、使用频率、增长速度各不相同。

收稿日期:2022-10-10; **修回日期**:2022-11-06。
基金项目:国家青年科学基金项目(61502196);吉林省自然科学基金项目(20200201290JC)。
作者简介:石晓栊(1988-),四川广元人,硕士,工程师,主要从事石油与天然气工程方向的研究。
通讯作者:彭 君(1981-),重庆人,博士,讲师,主要从事人工智能、机器学习、软件开发方法、软件体系结构方向的研究。
引用格式:石晓栊,赵统永,王耀忠,等.基于大数据应用的地质灾害数据存储策略[J].计算机测量与控制,2023,31(6):156-161.

地质灾害点数据：该数据是每年野外作业调查后形成的 word 文件，原始文件为 .doc、.docx 格式，内容以表格形式列出，包括每个灾害点包含编号、行政区、坐标位置、基本特征、照片、平剖面图、调查时间等内容，以文字、照片、图片等形式展现。

巡查巡测点数据：该数据也是野外作业调查后形成的 word 文件，原始文件为 .doc、.docx 格式，内容以表格形式列出，每个巡查巡测点包含编号、行政区、坐标位置、基本特征、照片、平剖面图、巡查巡测时间等内容。以文字、照片、图片等形式展现。

动态监测数据：该类数据由监测设备产生，数据格式以表格或数据库的形式存在，包括 Excel 表格，Access、SQL Server 数据库等。是各类型监测仪器采集的监测结果数据，包括雨量、应力、地表位移、土壤含水率、深部位移、视频等多种传感器类型的数据。

预警预报数据：该类数据是每年汛期发布的预警数据，包括预警地点、等级、预报词等信息。数据格式为文本格式、图片格式、shp 格式，主要以文本及图片形式存储。

报告类数据：该类数据包括地质灾害勘查设计报告、排查报告、风险评价报告以及地质防治相关报告、国家企业行业技术标准、监测报告等类别。数据格式为 PDF 文件。

地质/地理数据：该类数据包括地质环境基础数据（地层岩性、地质构造、区域地质、工程地质、水文地质、环境地质、矿产资源等）、基础地理数据（行政中心、行政边界、行政区、河流、湖泊、铁路、公路等）。数据类别分为 ArcGIS 格式、MapGIS 格式、遥感影像格式及 DEM 格式。

气象水文数据：该类数据包括了降雨数据、水文站数据、台风、雾况数据等。降雨数据、水文站数据格式为 shp 格式，台风、雾况数据格式为文本格式。降雨数据包括 24 小时、48 小时、72 小时实况降雨及预报降雨。

多媒体数据：该类数据包括 MP4 格式的视频数据及 JPG 格式的图片数据。

2　基于分布式的数据存储架构

2.1　硬件设备组织结构

针对地质灾害监测和预警所需要使用的数据特点，构建了传统关系型数据库与非关系型数据库相结合的分布式数据存储架构。为适应数据量不断增加、写入和读取并发量增大的情况，非关系型数据库均采用分布式模式进行构建，将数据存储和数据服务分散到多台数据库服务器上，以提高性能和均衡负载。对于预警平台所需使用的统计数据、筛查数据等，每次获取数据量少，但是需要频繁获取，查询方式灵活、复杂，比如基础相关信息表、部门相关信息表及系统相关信息表，此类数据采用传统的关系型数据库进行存储。对于 AI 模型及大数据分析所需要使用的源数据，一次数据的获取量庞大，获取的数据通常是大块、整体、连续的，查询索引方式简单，比如遥感图像、航拍图片、历史记录等，此类数据采用非关系型数据库进行存储。

如图 1 所示，分布式数据库的硬件设备可以分为两个一部：一部分用三台普通服务器构成主从结构，一台作为主服务器另外两台作为从服务器，用于部署非关系型数据库 HBase，满足大数据及人工智能等应用的需求；另一部分用一台高性能服务器部署关系型数据库 PostgreSQL，满足网页显示及实时查询等传统应用的需求。

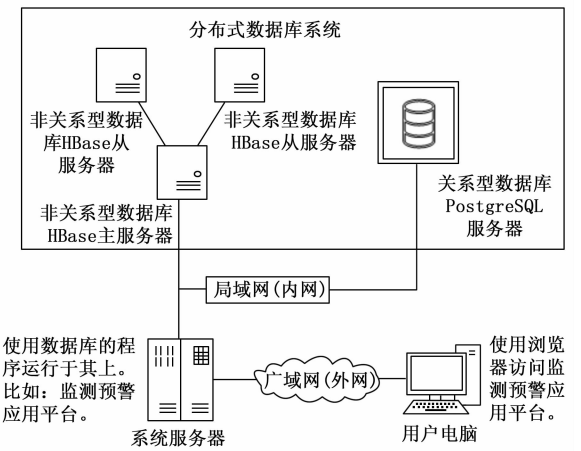

图 1　数据存储硬件设备组织结构

非关系型数据库 HBase 和关系型数据库 PostgreSQL 均部署在内部局域网中，仅供内网中的系统服务器进行访问，不对外提供访问接口。需要访问分布式数据库的应用都部署到系统服务器上，系统服务器上具有双网卡配置并配有防火墙，即可访问内网中的分布式数据库，又可为广域网中的用户提供对应的服务。

基于分布式存储的软件模块组织结构如图 2 所示。关系型数据库 PostgreSQL 被部署在一台安装了 CentOS 7.0 操作系统的虚拟主机上。非关系型数据库则是被部署到三台安装了 CentOS 7.0 操作系统的虚拟主机上。

PostgreSQL 服务器上安装 PostgreSQL 数据库服务端程序，在系统应用服务器上则安装 PostgreSQL 数据库驱动和 PostgreSQL 数据库客户端软件。

非关系型数据库是基于 Hadoop 生态系统进行构建的，包括了分布式文件系统（HDFS）、分布式列存数据库（HBase）、分布式数据仓库（Hive）、数据抽取转换器（Sqoop）等。底层数据存储使用 Hadoop 生态中的分布式文件系统 HDFS，HDFS 具有海量存储、高效访问和冗余备份能力。数据库系统采用 HBase，HBase 采用列式存储易于扩展，可以实时读写、随机访问超大规模数据集。Hive 提供类似 SQL 的查询功能。

HDFS 分别被部署到三台安装了 CentOS 7.0 的虚拟主机上，在其中一台虚拟主机上将 HDFS 配置为 HDFS 的名字节点，使其在该虚拟主机上启动 ResourceManager、NameNode 和 SecondaryNameNode 三个进程。在另外两台虚拟主机上将 HDFS 配置为 HDFS 的数据节点，使其分别在这两台虚拟主机上启动 NodeManager 和 DataNode 进程。系

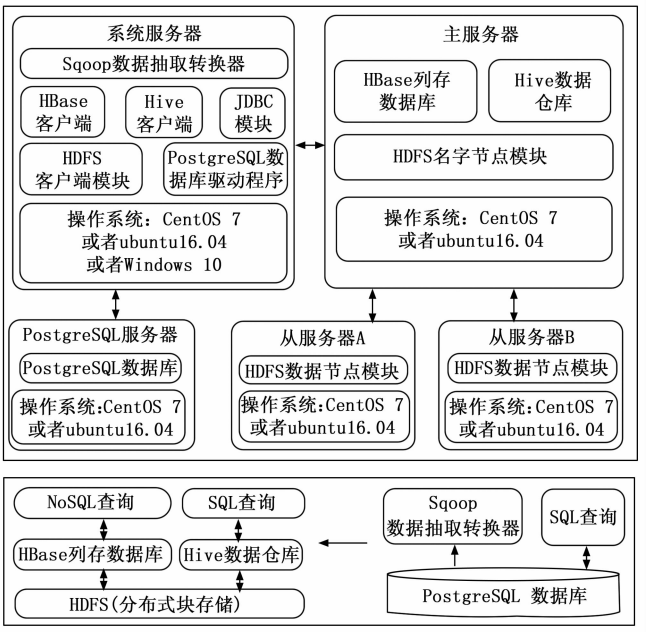

图 2 基于分布式存储的软件模块组织结构

统服务器作为 HDFS 分布式文件系统的使用终端，也需要在其上部署一套 HDFS 的客户端程序。

HBase 是部署在 HDFS 基础上的，在部署和运行 HBase 数据库之前需要完成 HDFS 的部署和启动。HBase 的也需要被部署到三台安装了 CentOS 7.0 的虚拟主机上。在部署了 HDFS 名字节点的主服务器上，将 HBase 部署为数据库服务端，使其在该虚拟主机上运行 HMaster、HQuorumPeer 和 ThriftServer 进程，为 HBase 数据库服务提供对外接口。在部署了 HDFS 数据节点的两台从服务器上，将 HBase 部署为数据存储端，使其分别在这两台虚拟主机上启动 HRegionServer 进程，负责 HBase 数据库中数据的实际存取。

本课题中构建的分布式数据库采用了关系型数据库与非关系型数据库相结合的分布式数据库架构。在对数据进行存储时，根据数据访问和使用的特点，将数据分别存储到不同的数据库中。对于统计数据、筛查数据等，每次获取数据量少，但是需要频繁获取，查询方式灵活、复杂，比如基础相关信息表、部门相关信息表及系统相关信息表，此类数据采用传统的关系型数据库进行存储。对于 AI 模型及大数据分析所需要使用的源数据，一次数据的获取量庞大，获取的数据通常是大块、整体、连续的，查询索引方式简单，比如遥感图像、航拍图片、历史记录等，此类数据采用非关系型数据库进行存储。

对于地图地质数据、地质灾害数据、巡查巡测数据，这些数据的原始数据格式是 word、pdf、shp 等复杂文本结构。对于这些数据的使用包括两个方面：一是根据关键信息索引对数据内容或者数据内容中蕴含的附属信息进行检索或查询。对于这样的数据应用，我们将数据的关键信息和附属信息抽取出来，在关系型数据库总设计相应的数据表来对其进行存储。另一种是直接使用数据的原始文件，基于原始文件进行数据的分析和数据挖掘。对于这种数据的应用，我们直接将数据的原始文件以字节流的方式存储到非关系型数据库中。

对于监测数据和预警预报数据，这类数据的生产频度很高，但是一次产生的数据量较小。对于这类数据的应用也是包括两种典型的情况：一是对监测数据和预警预报数据进行实时的查看和查询，每当有新的数据到来都需要进行获取。另一种情况是一次性批量的获取大量监测数据和预警预报数据的历史数据来进行数据的分析和预测等工作。因此对于监测数据和预警预报数据，一方面在关系型数据库中设计相应的数据库表来对实时数据、一定时间内的新到数据进行存储和更新，另一方面在非关系型数据库中也设计相应的数据库表，来对监测数据和预警预报数据的历史数据进行整体打包式存储。

3 基于数据粒度的存储模式

根据分布式数据库所采用的底层存储架构（HDFS）和数据库支撑架构（HBase）的特性[10-11]，以及需要存储数据的特点，为了获得良好的性能，首先从数据的粒度大小的角度来对数据进行分类和组织，对不同粒度的数据采取不同的存储方式。

根据数据的粒度大小将数据分为 3 个类别：小粒度数据，数据粒度小于 10 M 的数据；中粒度数据，数据粒度在 10 M 和 50 M 之间的数据；大粒度数据，数据粒度大于 50 M的数据。在所需要存储的数据中，小粒度数据主要包括：结构化信息、普通文本文件、地质灾害点数据、巡查巡测数据、预警预报数据、小图片视频数据。中粒度数据主要包括：气象数据、水文数据、管道本体数据、地质/地理数据、报告类数据。大粒度数据主要包括：遥感影像数据、动态监控视频、导出的超大文本、大型 Office 文件。当让这些不同种类的数据在大小上可能存在交叉的情况，在实际处理时，按照实际情况进行处理。

对于小粒度数据可以直接存储在 HBase 数据库中，将数据内容进行序列化之后作为某个列的值来进行存储。对于中粒度数据需要启动 HBase 的 MOB 特性后可存储在 HBase 数据库中，同样是将数据内容进行序列化之后作为某个列的值来进行存储，不同的是此时的列不再是普通列而是 MOB 列，对于 MOB 列 HBase 底层会进行分拆优化。对于大粒度数据不能将数据内容直接存储到 HBase 数据库中，如果直接存储会对性能造成较大的影响。根据 HDFS 的特性，适合存储较大的文件，HDFS 会对文件的存储自动进行分块优化。因此对于大粒度的数据，直接将数据以文件的形式写入到 HDFS 中，借助 HDFS 的读写优势来提高性能。同时为了方便地对数据进行查询和检索，在 HBase 中，以普通列的方式存储数据在 HDFS 中的文件索引，以便快速的获取文件。

除此之外，一些特别的情况，如某些数据历史性数据，数据量不大，写入频率也不高，写入后的访问频率也很低，只是作为历史性存档性质存在，在少数情况下会被查询。对于此类数据，也不直接存到 HBase 数据库中，而是按照大粒度数据处理的方式，直接写入文件系统，数据库中仅存储相关检索信息和文件索引。这样既节省了数据库的存储空间，又保证了高效、稳定的查询。另一种特殊情况，是对于实时动态监控数据，这类数据是由采集设备采集，定时、持续不断地发送给到服务器。此类数据粒度非常小，但是频率可能特别高，并且累计量大。对于这类数据的存储，按照小粒度数据的方式进行存储，在设计表的逻辑结构时考虑对其进行合并和压缩，以提高存储和读取性能。不同粒度数据存储方式如图 3 所示。

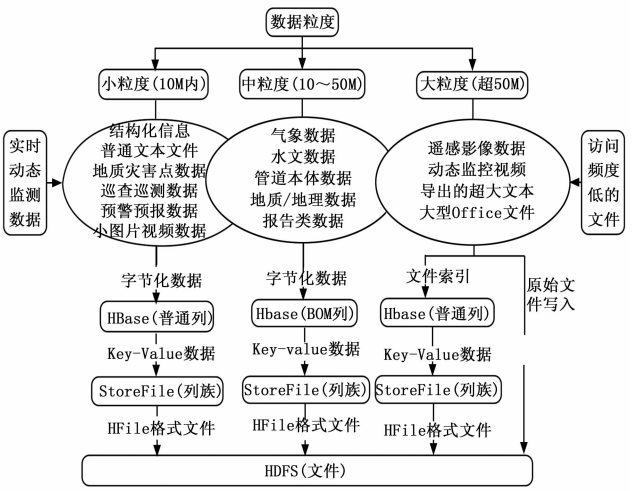

图 3　不同粒度数据存储方式

4　基于数据格式类别的访问模式

4.1　数据格式类别的划分

根据对需要存储数据的分析，数据包括了多种格式类型。根据导入数据库方式的不同，将这些不同的格式类型进一步归结为地理信息类、综合文档类、图片视频类、文本数据类和实时动态类，如图 4 所示。

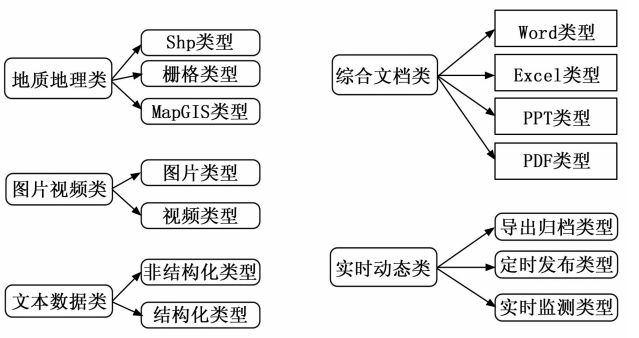

图 4　数据格式类别的划分

地理信息类主要包括了 shp 格式、栅格格式以及 MapGIS 格式。

综合文档类包括了常用的 Word、Excel、PPT、PDF 类型的文档，特别的对于规整的整齐的可以直接转换为 csv 类型的 Excel 文档可以不在此列，可以被当作格式化类型的文本数据进行处理。

图片视频类包括了图片和视频两大类格式。图片格式可以是 jpg、bmp、tiff 以及 raw 格式的原始信息图片。视频格式包括 mp4、mov、avi 以及监控系统产生的流式文件。

文本格式类包括结构化类型和非结构化类型两种。结构化类型文本是指文本中格式规范、统一的以特定的方式标记或分隔，能够直接转换为行列式结构的文件。非结构化类型文件则是指文本内容的归类或者切割不能直接完成，而是需要额外的处理和分析才能完成的文本类型。

实时动态类包括 3 种类型的数据：1）将大量历史数据或者累积数据按照特定的结构化格式归档、打包、导出的可解析数据；2）由某些机构、部门、或者公司综合整理统计后的某一方向、方面的集成化的数据。这类数据也具有良好的格式，可解析易处理易读取。不过包含的内容丰富，数据项目和条目都较多；3）由监控设备产生的实时监测数据，此类数据定时发送，数据内容单一、格式固定，但是频度高、具有持续性，单点数据量微小，累积数据量巨大。

4.2　数据的导入接口

对于 shp 类型的数据，这式地理信息系统中一种非常通用的数据，数据格式公开、清晰、明确，能够被多种地理信息软件所使用。因此对此格式类型的数据在进行数据库存储时，将其分成 3 个部分来进行存储。1）提取 shp 文件相关的属性信息，这些属性信息包括区域范围、数据内容、比例尺等描述说明信息，主要是用来对数据进行粗粒度检索。这些信息经过提取之后，存储到 HBase 数据库的列中；2）shp 数据原始文件，有的时候需要保存或者使用 shp 类型数据的原始文件，对于原始文件将其当作一个完整的对象进行处理；3）对于 shp 数据中的地理信息、附加信息等，则借助 GeoMesa 的 HBase 接口来进行入库处理。这种处理方式会读取分析 shp 文件的具体内容，并进行分析和分解，将分解后的数存储到 GeoMesa 构建的 5 张 HBase 数据表中。这 5 张数据表中构建了时空索引，为 GIS 引擎提供数据支撑。

对于栅格格式、MapGis 格式、WordF 格式、Excel 格式、PPT 格式、PDF 格式、图片格式、视频格式以及非结构化文本格式，这些格式数据的解析和分析，需要通过复杂的处理、专门的流程和专业的软件来完成，因此在数据库存储的时候，无法对具体的内容做更进一步的分解存储。对于此类数据，分成两部分进行存储：1）提取相关描述性、定位性信息作为查询、检索、分类指标直接存储到 HBase 的列中；2）原始数据作为完整的对象进行存储。对象存储的方式根据对象的大小来具体确定是直接保存在 HBase 内还是存放在 HDFS 文件系统中。此类数据经过分析处理工具和软件的处理后可以将处理后的内容重新入库。

对于结构化类型的文本数据、导出归档的监控数据、实时发布的综合监控数据。这些数据是格式化的文本文件，可以通过 HBase 提供的批量导入工具来导入到 HBase 数据库的表中。

对于实时监控类数据，为了能够提高存储效率，节省存储空间，为其专门设计了时序相关数据库表来对其进行存储。数据库表的行键、列簇和列通过专门的设计，具有整合和压缩数据的能力，能够快速高效地完成数据的存储。

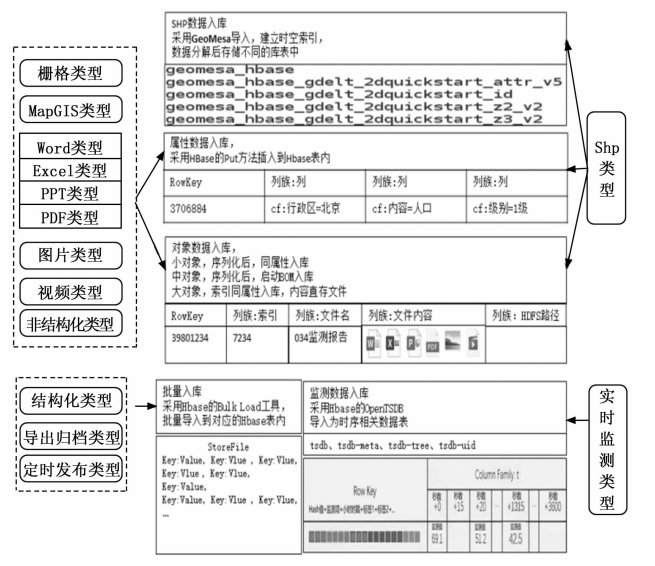

图 5 不同格式类别的导入模式

4.3 数据的使用和获取方法

数据存储到数据库之后，还需要能够被准确、快捷、便利地提取出来，用于查询、计算、分析等多种用途。对于不同类型的数据以及不同的用途，有不同的数据获取方式。

最基本的数据获取方式是通过 HBase 提供的 get 和 scan 方法，这个两个方法能够针对数据库中所有的数据进行获取，无论数据类型和存储方式。其中 get 方法用于获取指定的一行数据，scan 方法可以获得多行数据，相当于是多条件查询。这两个方法不仅可以在 HBase 命令行中被用户直接通过客户端来运行，也可以在 HBase 提供的 API 中被程度代码调用执行。

对于按照对象类型进行存储的数据，在采用 get 或 scan 进行查询的基础上，还可以通过 HBase 和 HDSF 提供的反序列化方法，将被存储的序列化的字节编码的数据还原为原始的对象类型，以便用户对存储数据的使用。

对于数据库中存储的 shp 格式的地理信息数据，除了可以通过基础查询和对象查询对数据进行简单的获取方式外，还可以利用 GeoMesaAPI[17] 直接对数据库中的地理信息数据进行深层次的查询分析和专业处理。GeoMesshforHBase 通过数据库表的优化设计，二维、三维时空索引的建立，丰富的查询、分析、计算接口，简化了用户对数据获取和处理的难度，降低里用户对数据使用的复杂度，扩展了 HBase 数据库中地理信息类型数据的应用范围和使用模式。

对于 HBase 中利用 OpenTSDB 时序数据库表进行存储的动态监控数据[18-21]，OpenTSDB 除了提供数据采集器帮助用户收集存储数据外，也提供了数据压缩、数据均衡性、聚合、插值、缩减采样等丰富的功能。此外提供了一个 web 页面供数据查询。

对于进行分析计算等处理之后的结果数据，也可以建立与原始数据相关联的数据库表来对其进行二次存储。

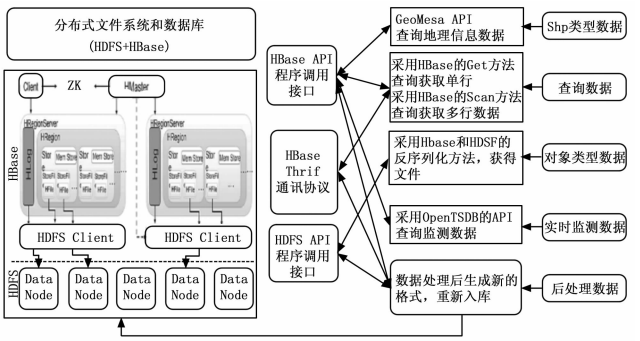

图 6 数据的使用和获取方法

5 地质灾害数据存储的实现

HBase 在数据存储的底层上采用了多 Region 的方式来实现，在为用户提供的应用接口层面上也是以表为基本单元进行组织的。在 HBase 的表中也具有行和列的概念，但与传统关系型数据库中的行和列具有不同的涵义。HBase 中的表并不像关系型数据库的表是整齐排列的，表中的列是按需出现的，某行数据只会将该行实际具有的列数据写入表中，该行中不存在的数据项不会在表中进行占位，因此 HBase 不会因为空值的存在而损失存储空间，能够很好地应对具有稀疏特征的数据存储。

HBase 中的行由行键来唯一确定，行键相当于表的索引，为了加快 HBase 的检索速度和存取效率，行键的设计起到关键作用。行键的设计首先要满足唯一性要求，并具有高效的索引能力；其次要满足散列性要求，能够避免数据写入时热点效应带来的影响；第三在满足散列性要求的同时还需要满足集中性要求，能够让用户在获取数据时一次性获取全部数据。这些要求使得行键的设计具有很好的灵活性和技巧性。

HBase 中的列族在数据库表设计完成后通常是固定不变的，会出现在每一行数据中。列族决定了表中数据在底层 HDFS 上的存储方式，属于同一个列族的数据在底层是存储在一起的。列族在数据库表设计完成后还可以进行添加，这为 HBase 提供了横向扩展能力。

HBase 中的列是依托于具体数据存在的，对于每行数据来说列都可以不相同，因此 HBase 中列的设计由具体的数据来决定。

地理信息类、图片视频类、文本数据类、气象水文类和实时动态类数据的通用逻辑结构如表 1～5 所示。

表 1　地理信息类型的数据表 Hbase 的逻辑结构

列族 / 列	cf							
行键类别标识＋范围标识＋内容标识＋文件标识	原始文件名	文件大小	文件创建时间	文件上传时间	文件标签	文件关键字	文件内容	正 HDFS 中文件路径

表 2　图片和视频类型的数据表 Hbase 的逻辑结构

列族 / 列	cf							
行键 Hash 值＋类别＋文件名	原始文件名	文件大小	文件创建时间	文件上传时间	文件标签	文件关键字	文件内容	在 HDFS 中文件路径

表 3　文本类型的气象和水文的数据表 Hbase 的逻辑结构

列族 / 列	cf								
行键站号逆序＋观测时间	气压	气温	地表温度	湿度	降雨量	风力	风向	水位	流量

表 4　气象水文站信息的数据表 Hbase 的逻辑结构

列族 / 列	cf							
行键站号	经度	纬度	省	市	县	地址	级别	…

表 5　动态监测数据的数据表 Hbase 的逻辑结构

列族 / 列	cf							
行键 Hash 值＋监测项目＋采集时间＋标签 1＋…＋标签 8	0	1	2	…	…	…	3 600	…

列名是秒的编号。列是可以动态增删的，在某一秒有数据则添加列，无则不添加。列值表示在某个时间的监测值。

6　结束语

本文根据地质灾害监测和预警等人工智能和大数据应用需求，对地质灾害监控系统、预测预报系统所需的数据进行了分析和统计。基于基于分布式文件系统（HDFS）和列式存储非关系型数据库（HBase）提出了地质灾害是数据存储策略。设计了基于数据粒度的优化存储模式，根据数据的大小自动采取直接存储、文件存储、压缩存储进行数据的入库保存。将地质灾害监测数据归并为地理信息类、综合文档类、图片视频类、文本数据类和实时动态类 5 个不同类别，针对每个类别的数据特性和使用方式，提出了不同的存储结构和访问接口，以获得最优的数据访问性能。

参考文献：

[1] ARAVINTH S S, SHANMUGAPRIYAA M S, SOWMYA M S, et al. An efficient HADOOP frameworks SQOOP and ambari for big data processing [J]. International Journal for Innovative Research in Science and Technology, 2015, 1 (10): 252－255.

[2] 李　孟，曹　晟，秦志光. 基于 Hadoop 的小文件存储优化方案 [J]. 电子科技大学学报，2016，45 (1)：141－145.

[3] 葛　微，罗圣美，周文辉，等. HiBase：一种基于分层式索引的高效 HBase 查询技术与系统 [J]. 计算机学报，2016，39 (1)：140－153.

[4] 崔　晨，郑林江，韩凤萍，等. 基于内存的 HBase 二级索引设计 [J]. 计算机应用，2018，38 (6)：1584－1590.

[5] TIANYI F, LI Y, ZONGMIN M. Storing and querying fuzzy RDF (S) in HBase databases [J]. International Journal of Intelligent Systems, 2020, 35 (4) : 751－780.

[6] 蒋叶林. 基于 HBase 数据库的时空大数据存储与索引研究 [D]. 昆明：昆明理工大学，2021.

[7] 邹　喆. 面向时空数据的 HBase 索引与查询技术研究 [D]. 重庆：重庆大学，2020.

[8] 王普刚. 基于 HBase 的工业日志系统设计与实现 [D]. 大连：大连理工大学，2016.

[9] 温　静. 基于 Hbase 的风电机组运行数据的存储与检索策略研究 [D]. 太原：中北大学，2020.

[10] JIZHE X, CHAOWEI Y, QINGQUAN L. Building a spatiotemporal index for earth observation big data [J]. International Journal of Applied Earth Observation and Geoinformation, 2018, 73: 245－252.

[11] RUIYUAN L, SIJIE R, JIE B, et al. Efficient path query processing over massive trajectories on the cloud [J]. IEEE Transactions on Big Data, 2020, 6 (1) : 66－79.

[12] 周笑天，冯　勇，陈益玲，等. 基于 Hadoop 的气象数据分布式存储技术研究 [J]. 信息技术，2022 (1)：68－74.

[13] 涂美义. 省级地质灾害应急服务架构及方法体系研究 [D]. 武汉：中国地质大学，2016.

[14] 张茂省，贾　俊，王　毅，等. 基于人工智能（AI）的地质灾害防控体系建设 [J]. 西北地质，2019，52 (2)：103－116.

[15] 王晓刚. 基于 HBase 与 HDFS 的洪灾数据存储与管理研究 [D]. 赣州：江西理工大学，2016.

[16] 王键键，王江燕. 大数据技术在城市洪涝灾害分析预警中的应用 [J]. 工程技术研究，2021，6 (20)：253－254.

[17] 孟鑫淼. 基于 HBase 的空间数据云存储研究 [D]. 郑州：河南大学，2016.

[18] 包文峰，郭慧斌，张志俊，等. 基于 OpenTSDB 数据库的测风塔管理系统开发研究 [J]. 风能，2019 (12)：70－75.

[19] 宋江健. 基于 OpenTSDB 和 OPC 的能耗数据采集存储技术研究 [J]. 福建电脑，2019，35 (1)：8－9.

[20] 杨　帆. 基于 opentsdb 的分布式实时监控方案 [J]. 福建电脑，2016，32 (11)：143，169.

[21] 单若琦. 一种基于 OpenTSDB 的海量实时数据存储系统 [D]. 广州：华南理工大学，2016.