

遥感影像地图服务即时发布技术及应用

田茂春 赖 杭

(珠江水利科学研究院, 广东 广州 510610)

摘 要 针对当前遥感影像大多停留在“从数据到数据”的阶段, 存在数据到服务应用上不足的问题, 基于分布式文件存储和分布式计算框架等技术, 实现了海量多源遥感影像数据的高效存储、快速检索、服务即时在线发布, 构建了影像切片即时发布服务平台。应用表明, 该技术可提供快捷便利的遥感影像数据服务, 为遥感数据的存储、管理以及应用提供有力支撑。

关键词 影像存储 地图切片 影像管理

中图分类号: TP7

文献标识码: A

文章编号: 1007-0745(2022)12-0027-03

随着科学技术的不断发展, 遥感技术和地理测绘技术也在不断的完善, 当前, 已经能获取到较高分辨率, 文件巨大的遥感影像数据。

针对影像数据比较大的特点, 运用金字塔模型的方式能够将不同分辨率的影像进行调整, 从而实现不同分辨率之间的有效浏览^[1-2]。

另外, 常规的串行方法通常适用于相对小规模的数据, 为对大数据量的遥感影像进行分析, 往往采用 MapReduce^[3-5], 但在针对高分遥感影像进行处理时, 所耗费的时间仍然相对较长, 并且海量遥感影像的可视化过程非常缓慢。

已有学者利用分布式计算框架来对遥感影像处理, 如存储、镶嵌、金字塔构建等方面做出了广泛研究^[6]。在遥感影像处理方面, 黄冬梅^[7]等提出了基于 Spark, 采用并行的方法来进行影像金字塔构建。在影像切片及元数据存储方面, 文献^[8]研究表明了可以利用 NOSQL 进行影像元数据或切片进行存储, 但大多是元数据用 NOSQL 存储, 瓦片数据用文件系统或者是多个数据库配合使用来达到目的, 此外, 对影响瓦片快速读取的索引策略并没有做更深入的研究。

在当前的技术中, 可以发现, 海量遥感管理与在线可视化浏览技术之间处于相互独立的状态, 为了有效地解决遥感技术信息存储的问题需要结合遥感数据发布的需求进行调整。

本文基于分布式文件存储、分布式计算框架等技术, 开展了遥感影像数据的分布式存储、高效检索和地图服务即时发布等关键技术研究, 构建影像即时切片服务发布平台, 能够使大量堆积的数据得到有效管理和利用, 以满足项目影像服务应用需求。

1 相关技术

1.1 Hadoop

Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构。开发者可在不了解分布式底层细节的情况下, 开发分布式程序, 充分利用集群的威力进行高速运算和存储。Hadoop 实现了一个分布式文件系统 (Distributed File System), 其中一个组件是 HDFS。随着 Hadoop 不断发展, 存储结构已不限于最初的 HDFS(大文件存储), 而像 HBASE 分布式列数据库也得到了大量发展和应用。

而在计算引擎方面, 除了 MapReduce 这种适合离线计算的引擎外, 又扩展了适合内存计算的 Spark 计算框架以及实时流计算框架 Storm。此外, Hadoop 生态圈还包含了日志收集器 Flume、脚本 Pig、资源管理器 Yarn 等组件。

1.2 Spark

Apache Spark 是当前针对大数据处理设计的计算引擎, Spark 对大规模数据的快速处理优势使其在海量影像计算处理方面有着广泛应用基础。

1.3 MongoDB

MongoDB 是 10gen 公司基于 C 语言研发的一个开源的基于分布式文件存储的 NoSQL 数据库系统。相比关系数据库, 易用性更强, 功能非常丰富, 扩展了比如地理信息查询、储存等功能。另外, 查询语言非常强大也是 MongoDB 的一大特点。

上述 MongoDB 的特点使其非常适用于数据量大, 查询性能较高的业务, 特别是针对多元遥感影像元数据管理。

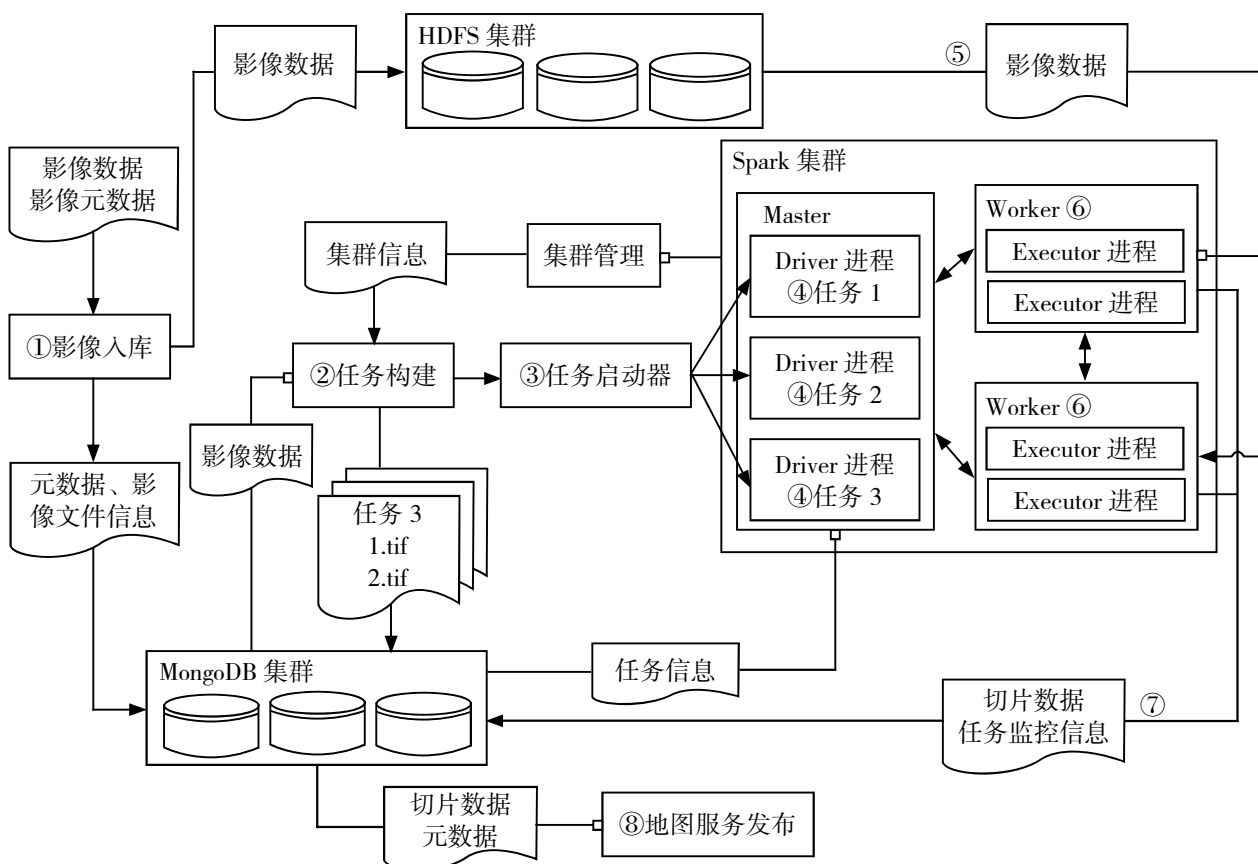


图 1 影像即时切片服务发布流程图

2 关键技术

2.1 遥感影像元数据存储与检索

遥感影像来源多样, 描述其数据的元数据往往也各不相同, 传统的 xml 文件存储方式读取解析繁琐, 不易于应用系统集成使用, 而如果以关系型数据库存储, 由于各元数据属性差别较大, 为同时满足不同源元数据存储, 关系表是所有类型遥感影像元数据的并集, 存储时存在很多字段为空情况, 造成存储资源浪费。而 NOSQL 数据库的代表——分布式文件数据库 MongoDB, 以类 JSON 方式进行数据存储, 可随时对存储结构进行修改, 特别适合于多源遥感影像元数据存储。

本文设计了用于建立多星源遥感影像关键属性(时相、产品名称、分辨率等)到遥感影像元数据表的映射表, 入库时通过该模板表将 xml 中元数据属性映射入库。同时, 设计了用于存储各种星源的元数据, 入库时根据元数据关键属性模板表, 将解析的 xml 文档元数据主要属性映射成本表中关键属性, xml 文档元数据剩余属性按其 (tag,value) 也存入该表的影像元数据表。

在元数据入库及显示方面, 根据遥感影像检索以

及元数据查询显示需求, 采用设计入库模板、显示配置模板的方式支持任意类型的多源遥感影像入库及查询显示。在元数据索引方面, 采取建立地理空间索引 + 多个关键属性的复合索引键, 如地理范围、时相、观测时间、云量、分辨率、传感器类型等来实现遥感影像快速检索。

2.2 基于 HDFS 的遥感影像分布式存取技术

遥感影像分辨率高、文件大, 传统的文件系统存储方式管理不便、安全性低, 且单机存储不适合在高并发访问下使用。

结合遥感影像数据的特点, 一般会选择 HDFS 作为分布式文件管理的数据库, 其能够支持数百个储存节点, 另外, 可以使用添加节点的方式来扩充储存, 而且可以高吞吐访问应用程序数据, 适合那些有海量数据集的应用。利用分布式文件系统 HDFS 将遥感影像大文件切成几十或上百兆大小的数据块, 每块分散在分布式集群各数据点并以复制集的方式存储, 提高了数据存取的并行性以及数据安全性。另外, 分布式计算框架“将计算移动到数据”的特点, 也充分利用

了分布式文件系统分散存储的优势。

遥感影像在 HDFS 上的存储采取多副本方式,其中每一个数据块中有多个副本,并且储存于不同的服务器之上,当前,工业界会使用三副本的方式,通过这样的形式,能够有效地保证数据的使用不会丢失,能够恢复相应的数据块,并由此保证提供数据的高效性,同时,读数据时可以提高读的效率。

2.3 基于 MongoDB 的切片存储

MongoDB 可以存储结构非常复杂的数据类型,尤其是非结构化数据,其充分利用内存的方式得到了比关系型数据库 Mysql、Oracle 大数倍的查询吞吐率,在项目中研究了在 MongoDB 数据库上进行“小文件”式的影像瓦片存储。切片索引键设计和切片数据,设计 MongoDB 切片存储文档结构。MongoDB 插入文档时自动生成的主键, TID 为切片索引键,数据类型为字符串, IMG 是切片数据,为二进制数组 Binary 形式。

2.4 基于 Spark 的遥感影像分布式并行瓦片构建方法

通过利用金字塔模型,根据不同分辨率进行影像的调整,这对于数据库的储存和运行效率有较高的要求,然而,常规的串行计算方法仅仅适用于内存较小,规模较小的遥感数据。当前所利用 Map Reduce 的分布式计算框架方式能大大提高金字塔构建效率的方法,但是在涉及海量、多元以及高分遥感影像时,所耗时间仍然很长。而分布式计算框架 Spark 由于将计算中间结果存储于内存,大量减少了中间结果存取的时间,在分布式计算领域得到了广泛应用。在研究 Spark 计算框架转换、行动算子基础上,将瓦片任务分解成并行执行的转换算子和行动算子算法,实现影像瓦片的并行构建。在影像瓦片输出方面,由于影像瓦片数据具有单个瓦片数据量小、瓦片数量巨大的特点,通过研究 Spark 分区特点,设计将每个分区多个瓦片合并为一个批次输出至数据库来提高并行性。

2.5 影像即时切片服务发布平台

在前述关键技术研究基础上,将 SSM 框架与 Spark 分布式计算框架进行集成,梳理遥感影像地图服务应用方式,构建海量遥感影像从存储、瓦片并行生成、地图服务即时发布、地图服务即时共享使用的影像即时切片服务发布平台,为各个项目提供影像数据服务。影像即时共享发布流程如图 1 所示。

3 应用案例

影像即时切片服务发布平台已成功应用于河湖长制、河湖岸线监管、水土保持监管等依赖大量遥感影

像数据的项目中,为项目中遥感影像的管理、查找及快速查看提供了技术及平台支撑,进一步提高了遥感数据的使用效率。利用平台可进行多源海量的遥感影像的并行切片与生成任务在线提交,同时,可以快速查看原始遥感影像数据及元数据信息,方便用户筛选数据质量最优的遥感影像用于遥感解译及后期的数据应用业务。

4 结论

本文利用分布式文件系统、分布式数据库、分布式计算等技术,对多源遥感影像数据进行高效存储,完成了遥感影像快速构建金字塔、瓦片的并行计算与生成;设计了影像元数据、影像瓦片的索引算法,实现影像高效检索、瓦片快速读取,构建了影像切片即时发布服务平台,为更有序、更高效管理遥感数据,快速提供遥感地图服务及可视化显示,实现“数据到服务”提供了一体化技术方案。

参考文献:

- [1] 程承旗,张恩东,万元菟,等.遥感影像剖分金字塔研究[J].地理与地理信息科学,2010,26(01):19-23.
- [2] 丁凯孟,朱长青.基于金字塔模型的遥感影像感知哈希认证算法[J].地理与地理信息科学,2015,31(02):16-19,F0002.
- [3] 刘义,陈萃,景宁,等.利用 MapReduce 进行批量遥感影像瓦片金字塔构建[J].武汉大学学报:信息科学版,2013(03):278-282.
- [4] 刘小利,徐攀登,朱国宾,等.结合 MapReduce 和 HBase 的遥感图像并行分布式查询[J].地理与地理信息科学,2014,30(05):26-28,32.
- [5] 沈盛戡,刘哲,张平仓,等.基于 MapReduce 的高分辨率遥感影像特征提取方法[J].长江科学院院报,2014,31(02):91-96.
- [6] 黄震,钱育蓉,范迎迎,等.Spark 下遥感大数据特征提取的加速策略[J].计算机工程与设计,2017,38(12):3279-3283.
- [7] 黄冬梅,杨雨浩,梅海彬,等.面向 Spark 的遥感影像金字塔模型的并行构建方法[J].计算机应用与软件,2017,34(05):175-181.
- [8] 李宏志,李菟兰.基于 MongoDB 集群的遥感数据存储方法研究[J].山东师范大学学报(自然科学版),2018,33(03):293-301.