

基于大数据平台的图像数据库架构的设计与实现

王莉, 张勇

(宿州职业技术学院计算机信息系, 安徽 宿州 234101)

摘要: 图像识别及其图像修复技术在目前的数字化领域应用非常广泛, 在金融、安防领域尤为突出。因此, 从经济和实用性角度出发基于大数据的图像修复技术有着广阔的应用前景。但大数据环境下复杂的数据存储是数据图像修复的关键, 如何建立一个高效、安全、高容量的基于大数据平台的数据库, 是数据库建立需要考虑的关键问题。下文从数据库架构的构成、应用目的、数据量大小、访问量、安全要求等各个角度研究和讨论图像数据库的建立, 力求使数据库建立在安全、合理、流畅的数据库架构之上, 达到最优的数据库架构模型。

关键词: 图像识别; 图像处理; 大数据; 架构

中图分类号: TP392 **文献标识码:** A

Design and Implementation of Image Database Architecture Based on Big Data Platform

WANG Li, ZHANG Yong

(Department of Computer Information, Suzhou Vocational and Technological College, Suzhou 234101, China)

Abstract: Image recognition and image restoration technology are widely used in the current digital fields, especially in the fields of finance and security. Therefore, from the perspective of economy and practicability, image restoration technology based on big data has broad application prospects. However, complex data storage in big data environment is the key to data image restoration. How to build an efficient, safe and high-capacity database based on big data platform is the key issue to be considered in database establishment. In this paper, the establishment of image database is studied and discussed from the aspects of the structure of database, application purpose, data volume, access volume, security requirements and so on, in order to build the database based on a safe, reasonable and smooth database architecture and to achieve the optimal database architecture model.

Keywords: image recognition; image processing; big data; architecture

1 引言(Introduction)

数据库作为数字信息的载体在数字生活中起着极其重要的作用, 无论是动态网站设计、数据挖掘、大数据存储等各个领域都很重要。数字图像作为数字信息的表现形式有着其他数字信息不可比拟的特点, 存量大、格式复杂、程序处理困难等都是其固有的属性。因此, 基于大数据平台的图像数据库架构的建设在整个平台建设中显得非常重要。

2 图像数据的分类及特点(Classification and characteristics of image data)

随着数码相机、扫描仪、视屏摄录设备的广泛应用, 越来越多的数字图像数据大量产生, 这些数据图像因格和采集

设备等的不同有着很大的差别, 不同的格式及设备采集的图像特点也不相同。

2.1 位图数据

位图数据由若干点阵构成, 每个点称为一个像素。文件大小与像素密集度有关。当文件太大时处理速就会变慢, 但点阵所表达的图像色彩丰富、逼真, 品质较高。

2.2 JPEG图像数据

JPEG是由国际标准组织(ISO)和国际电话电报咨询委员会(CCITT)为静态图像所创建的第一个国际数字图像压缩标准, 其不仅可以提供有损压缩也可以提供无损压缩。比传统的图像压缩方式有不可比拟的优越性。其特点如下: 一次将图像

由左到右、由上到下顺序处理；当图像传输的时间较长时，可将图像分数次处理，以从模糊到清晰的方式来传送图像^[1]；其压缩出来的图像可以在较低分辨率的设备上显示出高分辨率的效果。

2.3 TIFF标记图像文件格式

TIFF支持可选压缩、可扩展格式许多可选功能，TIFF是广泛支持的格式，尤其是在Macintosh计算机和基于Windows的计算机之间^[2]。以任何颜色深度存储单个光栅图像，其缺点是TIFF在Web浏览器得不到支持，而且TIFF之间的文件兼容性也存在问题。

2.4 PNG格式

PNG(Portable Network Graphics)是近几年来网络上比较流行的图像格式。其优点是目目前保证最不失真的格式，存贮形式丰富，兼有GIF和JPG的色彩模式的特点；其缺点是不支持动画应用效果。

3 大数据数据库设计与优化(Design and optimization of big data database)

3.1 数据库架构的设计

数据库架构是数据资源布置方式的统称，一个合理的架构可以提高数据的安全性和数据访问的速度。目前提供大数据访问的平台提供商，广泛采用B/S结构的数据库前端平台搭建模式(图1)。B/S结构具有使用方便、成本低、适应强等特点，但这种基于动态网站建设技术的数据平台前端建立模式最大的缺点就是安全性不能得到很好的保障^[3]。因此，在大数据平台下数据库的架构设计必须采用相关的措施提高数据库的安全性。

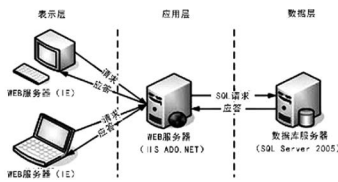


图1 B/S 结构图

Fig.1 B/S structure

为了提高数据库平台的安全性，在数据库核心层的外围需采用多层控制机制，通常从平台的使用、管理、控制、维护等方面数据库架构可由：产品层、接入层、网络层、中间件、存储层、运维服务等几个层级(图2)。

各层的具体作用如下：

产品层：该层是数据库架构的硬件平台，平台设备由服务商提供。主要负责整个架构软硬件设备的管理和使用。对于性能的要求与平台硬件的构成有关，基本要求是平台必须具备对整体架构安全性、实用性、操控性的管理^[4,5]。

接入层：接入层负责对外提供数据服务，对于大数据平台，如果要提高服务品质，良好的接入层设计可以使数据得

到高效的应用。大数据平台要面向众多的使用对象，接入层是每个使用对象必须经过的访问节点。接入层要提供有线、无线、云上数据等各种接入服务。如果从数据的安全性考虑，对众多的接入请求，接口要建立足够的安全保障机制。但如果复杂的安全保障机制建立在接口层，会使接入速度降低、服务品质下降^[6]。从大数据的使用初衷考虑在服务品质 and 安全性这两方面权衡，在接口层一般对安全性不做硬性要求，从而保证提供较高的服务品质。

中间件：大数据的特点就是数据量大而复杂，如何使用户快速、便捷地从海量数据获取所需信息，是架构设计必须考虑的。目前，多采用多语言支持、SQL二级索引等中间件来提高访问效率。另外，在中间件的下层多采用HBase面向列的开源数据库提供数据库技术支持。

存储层：该层作为数据的存储空间的管理层是整个数据库架构的核心，目前采用Ali-HDFS、共享存储OSS等管理数据资源，为数据保存提供高效组织方式^[7]。

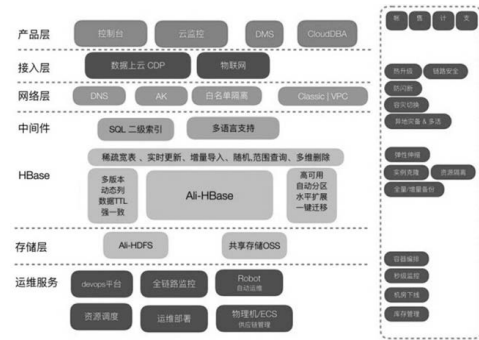


图2 数据库架构层级

Fig.2 Database architecture hierarchy

运维服务：大数据架构的使用对象众多，使用方式和习惯差异很大，平台架构运行后会有很多问题需要进一步服务。只有在不断的维护和改进中平台的功能才能更加完善^[8]。现在平台维护主要涉及全链路监控、自动运维、资源调度、运维部署、物理机等几个方面。

3.2 安全数据库的建立

为了提高大数据平台的服务品质，在接口层对于安全性没有采用过多的设计。但数据的安全还是大数据系统必须考虑的问题。因此，除了在产品层对于安全性的总体硬件部署之外，在数据库的结构设计上安全性的本质要求必须充分体现(图3)。

数据库是大数据系统的核心，因此安全层次的部署也相当复杂。如图3所示，可以采用专用安全数据库系统对核心数据库进行保护，数据库安全管理层结构主要构成及功能如下：

统一安全管理：主要功能负责整个数据库安全管理层的高层管理，具体有配置管理、安全审计、操作日志管理、用

户管理、权限管理、日志分析等功能。

监控管理子系统：包括警告管理、性能管理、配置管理，在警告管理模块主要实现警告过滤、警告分析、预处理、格式化等功能；性能管理模块主要包括数据稽查、阈值对比、数据汇总、预处理等功能；配置管理主要包括数据整合、异常处理、数据审核等功能^[9]。

安全数据库：核心数据库是大数据平台提供服务的信息载体，有着极其严格的配置管理控制。在安全控制及信息过滤的过程中也会需要大量的数据需要建立数据库进行存储，但从数据结构及安全配置管理的角度去考虑，这些存储的数据不需要供给大数据客户使用，因此需建立专用的安全数据库进行存储。这类数据库主要有警告数据库、性能数据库、配置数据库等。

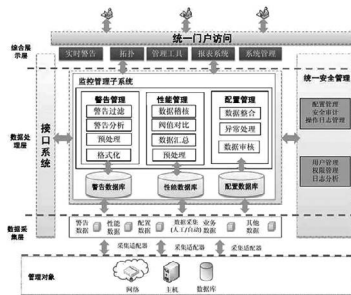


图3 数据库安全管理层次结构

Fig.3 Hierarchical structure of database security management

3.3 核心数据库的构成

当数据库的安全控制功能分离出去后，核心数据库的设计主要涉及大数据采集、大数据清理、大数据标准化、大数据结构化几个方面^[4](图4)。



图4 核心数据库结构

Fig.4 Core database structure

大数据采集：包含业务数据汇集系统、用户行为数据采集系统、大数据爬虫采集系统。

大数据清洗：包含业务数据清洗系统、用户行为数据清洗系统、互联网公开数据清洗系统。

大数据标准化：包含用户多ID匹配系统、商品多ID匹配系统。

大数据结构化：包含用户标签管理系统、商品标签管理系统。

图形图像处理是目前需求比较大的业务，在安防、医

学、考古领域都有涉及。尤其是图像修复更有着广泛的应用市场，人脸识别、考古修复等都在应用。因此基于大数据的图像修复系统中，图像数据库的建立尤为重要。图像数据具有文件大、种类多、存储结构复杂等特点。在建立图像数据库时所得到的数据同样存在着不可预知的数据^[10]。为了更精确的定位有用的存储信息，核心数据库建立时必须按照大数据采集一大数据清洗一大数据标准化一大数据结构化的步骤对数据进行优化，以便客户在使用时获得更为精确的图像数据。

4 结论(Conclusion)

基于大数据的图形图像处理是大数据的主要应用方向,良好的数据库架构可以提高处理的速度,减少误差,最大限度地提高工作效率。按文中所述，构建合理安全的大数据数据库架构是大数据平台建设必须遵循的规律，也是大数据应用发展的必经之路。

参考文献(References)

- [1] Trevor H.Booth.Species distribution modelling tools and databases to assist managing forests under climate change[J]. Forest Ecology and Management,2018(04):196-202.
- [2] Nagori N P,MalodeV.Communication Interface for Deaf-Mute People using Microsoft Kinect[J].International Conference on Automatic Control and Dynamic Optimization Techniques,2017(03):1-5.
- [3] Gu B B,Li Z X,Zhang X L,et al.The interaction between schema matching and record matching in data integration[J].IEEE Transactions on Knowledge and Data Engineering,2016(05):187-192.
- [4] 傅颖勋,罗圣美,舒继武.安全云存储系统与关键技术综述[J].计算机研究与发展,2013(01):136-145.
- [5] 张勇.数字图像大数据中心的研究与实现[J].淮阴工学院学报,2017(01):30-33.
- [6] 陈云亮.分级存储系统中基于进化算法的数据管理与保护关键技术研究[D].华中科技大学,2013:37-42.
- [7] 任崇广.面向海量数据处理领域的云计算及其关键技术研究[D].南京理工大学,2013:23-26.
- [8] 肖辉辉,段艳明.基于改进花授粉算法的移动机器人路径规划研究[J].软件导刊,2018(11):22-25.
- [9] 谢光.数据库大数据量存储结构的探索[J].通讯世界,2017(11):29-30.
- [10] 张玉英.目前信息技术背景下的数据库安全技术[J].电子技术与软件工程,2017(13):261-263.

作者简介:

王 莉(1979-),女,硕士,讲师.研究领域:数据库设计.
张 勇(1977-),男,硕士,教授.研究领域:图形图像处理.