Computer Science and Application 计算机科学与应用, 2017, 7(9), 879-885
Published Online September 2017 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2017.79100

Hans汉斯

Library Application of the Big Data: Opportunity and Challenge

Qingli Yu

Zhejiang Sci-Tech University, Hangzhou Zhejiang
Email: yuqingli.zju@qq.com

Received: Sep. 2nd, 2017; accepted: Sep. 14th, 2017; published: Sep. 21st, 2017

Abstract

With the advent of the cloud computing era, big data also attracted more and more attention. As a new important resource, big data is no longer just the data itself. It has changed the traditional business model and has also brought great opportunities and challenges for the information service industry. This paper explores the library framework under the big data environment from three stages: data analysis and integration, big data processing and result display. In the era of big data, library data processing and services will have a major change.

Keywords

Big Data, Library, Cloud Computing

大数据在图书馆的应用：机遇与挑战

俞晴里

浙江理工大学图书馆，浙江 杭州
Email: yuqingli.zju@qq.com

收稿日期：2017年9月2日；录用日期：2017年9月14日；发布日期：2017年9月21日

摘　要

云计算时代的大数据已受到越来越多的关注。作为一种新兴的重要资源，大数据这个含义已不单纯的是数据本身，它改变了传统的商业模式，也给信息服务业带来了机遇和挑战。本文从数据的分析与整合、大数据处理以及结果显示三个阶段描述大数据环境下的图书馆。在大数据时代，图书馆的数据处理能力和服务方式将会发生重大的改变。

文章引用: 俞晴里. 大数据在图书馆的应用: 机遇与挑战[J]. 计算机科学与应用, 2017, 7(9): 879-885.
DOI: 10.12677/csa.2017.79100

关键词

大数据，图书馆，云计算

1. 大数据的背景介绍

近几年，大数据这个名词，在各行各业引发了热议，随着科学技术的发展和进步，各式各样的记录信息行为的类型数据日益增多，这些数据和传统存储在库中的结构化数据统称为大数据(Big Data)。近年来，“大数据”这个词汇变得非常火爆，目前已经发展成为了国家和政府层面的发展战略。早在二零一一年，麦肯锡公司发布了一篇报告，在该报告中，第一次提到了大数据这个名词的概念，同时在该篇报告中还指出数据的重要性[1]。大数据这个含义已不单纯是数据本身，作为一种新兴的重要资源，它已开始改变传统的商业模式，为信息服务行业提供了新的途径。

2. 大数据对图书馆的影响

21 世纪，以数据为主导时代已经到来，它不但影响着国家的安全稳定发展，同时也推动着产业升级、经济进步和新产业的诞生，大数据影响到社会各个领域，当然大数据也对信息服务的中心——图书馆带来一定的影响和挑战。资源数字化、管理知识化、服务网络化是当今大数据环境下图书馆的主要特点，对我们的生活产生了重要影响。

2.1. 大数据下的图书馆对数据分析和挖掘潜在的价值

图书馆在质量服务和革新等方面都离不开大数据，大数据的作用在图书馆中越来越大。在图书馆中，数据种类十分复杂繁多，伴随着信息资源量的迅速增长，以及用户服务需求的不断增高，图书馆需要各式人才的信息。图书馆为了探索新的服务方式，还要对读者身份、借阅书目和记录等数据进行异构处理，因而在当前形势下，图书馆对数据的处理分析和对潜在信息价值的挖掘迫在眉睫。

2.2. 大数据对图书馆的数据存贮产生考验

当前图书馆主要由业务数据和应用数据等组成，应用数据往往需要更大的存储空间，其数据的安全性要求相对较低[2]。图书馆在对设备管理和维护的人力物力投入比较大，增加了经济成本。由于图书馆丰富的数据和信息以及各式服务，使得图书馆数据资源已经达到使用 TB 级或者 PB 级，数据的安全性无法得到完全保障，因此传统数据存储方式将面临淘汰的风险，而且图书馆的传统存储设备经常会遇到兼容性问题，这使得在不同的存储设备之间在线扩容不能很好的实现。

2.3. 大数据给图书馆信息服务带来影响

在大数据时代下，图书馆发生了许多改变，但同时也出现了一系列新的安全问题。随着目前网络化服务越来越方便，读者的信息将会遗留很多。由于数据具有很强的关联性和累积性，较多的个人行为信息的集中就可能加大用户隐私被暴露的可能[3]。此外，图书馆安全防护措施将会受到数据容量的影响，一般的手段会耗费大量时间，难以保障数据的安全需求，而且安全防护手段的更新升级速度非常缓慢，

这就使得大数据的安全性能更容易受到威胁。现阶段，信息咨询服务引入了云计算等技术，利用云计算技术解决了海量数据存储与处理的问题。传统的信息咨询组织能力、分析能力和储存能力有限，大数据时代的来临，使信息咨询服务创新看到了希望。

3. 图书馆实施大数据战略的必要性与可行性

3.1. 必要性

首先，大数据时代的图书馆，其资源数据化、服务信息化能力直接决定了未来图书馆的服务模式，即转向网络化和数字化。因此，将纸质文献进行数据化处理并保存到数据库，已成为图书馆资源建设的重要工作。这种结构化和非结构化的数据信息极大丰富了图书馆的文献资源，为新时期图书馆的数字化虚拟服务提供了有效途径。图书馆实施大数据战略，主要包括对文献资源的格式转换、储存、分析整合，从而向读者用户精准的数据分类和推送，增强读者用户对其的粘性。

3.2. 可行性

图书馆实施大数据战略需要得到技术支持和政府支持。对技术支持而言，需要将纸质文献进行数据化处理和数据库建设，因而需要高性能的资源云和海量的数据传输。国内大多数图书馆已经具备了高速的网络环境和数据处理条件。对政府支持而言，需要涉及多个政府部门参与图书馆的数据建设和信息化服务，需要政府在资金、政策上能够有相应的支持，如政府牵头将档案馆的纸质资料数据化图书馆。

4. 大数据环境下的图书馆框架

大数据环境下图书馆的框架主要分为以下三个阶段：数据的分析及整合、大数据处理和结果显示，如图 1 所示。其中数据分析及整合是整个框架的核心，大数据处理是技术层面的要求，结果显示为输出。

4.1. 数据的分析及整合

图书馆对数据的分析与整合是整个系统框架的核心，是将分散在不同类型平台和介质的数据实行挑选整理，剔除重复无效的资源数据，将同一类的数据信息进行分类整合，补充不完整的数据，该过程是整个提高数据质量的前期准备阶段，它和云数据储存平台(中转平台)有密不可分的关系(如图 2 所示)。

大数据环境下，数据载体是数据库建设的临时文件夹，至关重要，它用来实现数据的存储与信息交流[4]。具体来说，大数据环境下，图书馆对于数据的整合和补充为第二阶段数据处理与第三阶段结果显示做了铺垫，是整个框架构建的核心环节。传统的数据库往往由于数据信息的杂乱无章造成数据库的准确率和利用率不高，且整个信息资源十分分散，这会大大增加图书馆任务量，难以精确定位用户读者的需求，是数据服务体系上目前需要解决的问题，而大数据战略下的数据分析与整合大大的为这个问题找到了解决办法，更加方便了图书馆对信息资源的管理。

4.2. 大数据处理

图书馆大数据处理的过程主要建立在数据资源的分析及整合基础之上，收集大量的数据信息，并且通过信息之间的相互关联建立发散性的联系与链接。在大数据处理环节，图书馆需要依靠特殊的软件和技术，例如 Hadoop 系统和跨库检索等[5]，不但要对数据信息进行宏观的整理链接，还要对不同层面的数据信息进行关联，实现全方位的数据分析处理。

Hadoop 系统是一种开源软件，基于 Java 语言构建的 Hadoop 框架(如图 3 所示)，从根本上来说是一种分布式处理的大数据平台，包括软件和众多子项目。基于 Hadoop 的数据分析系统拥有很多优点：1) 稳定性高；2) 运算速度快、吞吐量大；3) 使用方便；4) 经济实惠。数据仓库和语境搜索也是大数据环境下所

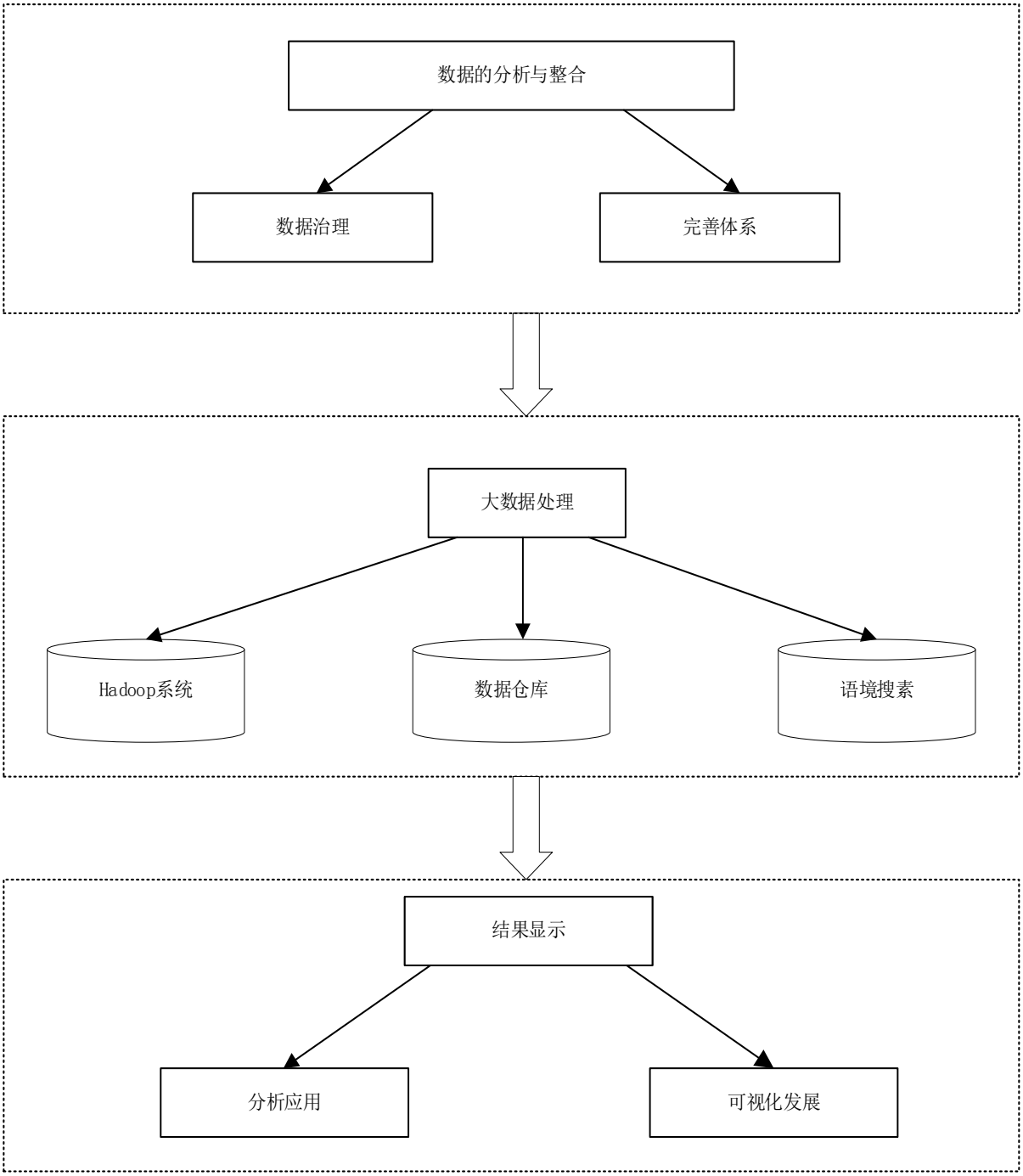

Figure 1. Library organ chart in the big data environment
图 1. 大数据环境下图书馆的框架

普遍采用的处理软件，语境搜索软件可以经济高效分析 PB 级结构化与非结构数据[6]，可以增强各节点之间的显性或者隐性联系，提高数据库的完整性，这使得图书馆的服务质量体系得到大大的提升。Hadoop 是由许多元素组成的，其中最底部是 Hadoop Distributed File System (HDFS)，它是存储 Hadoop 集群中所有存储节点上的文件。我们通过对 Hadoop 分布式计算平台最核心的分布式文件系统 HDFS、MapReduce 的处理可以实现对数字图书馆的分布式部署。

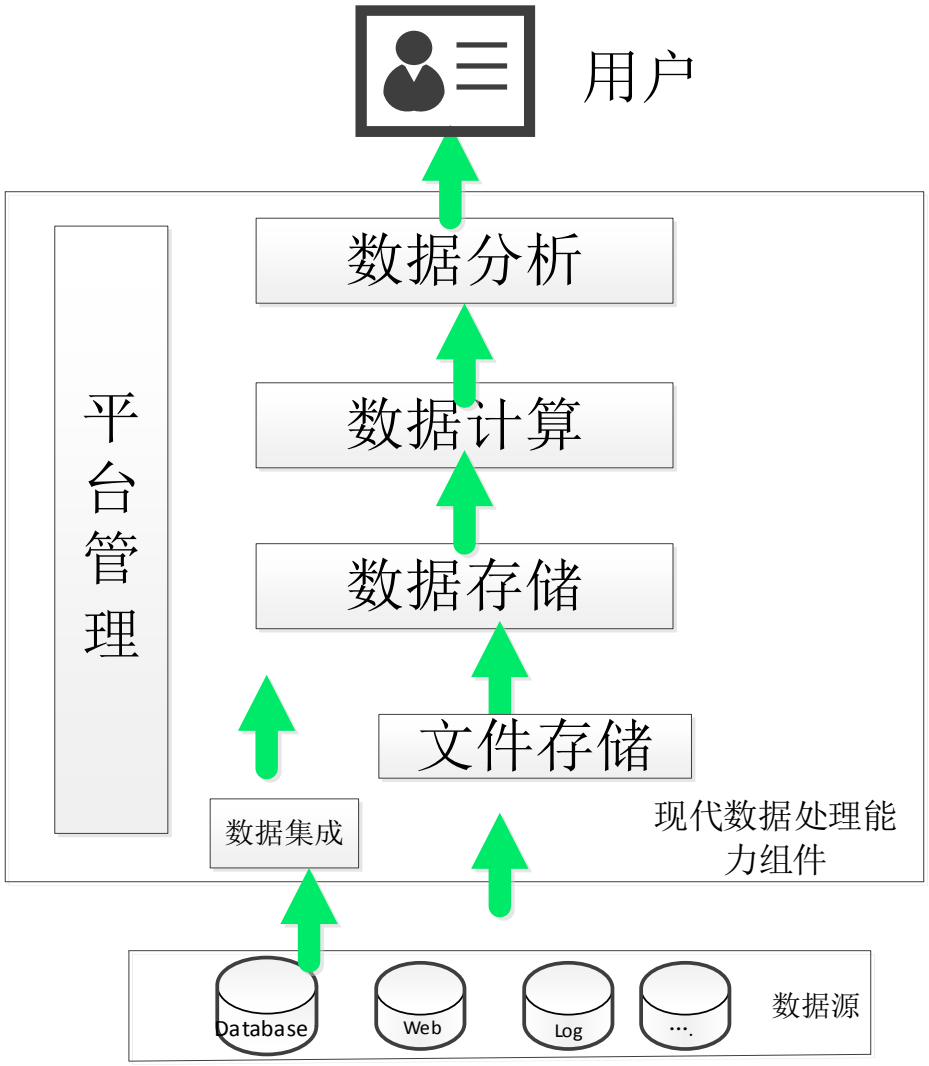

Figure 2. The framework of cloud data processing platform
图 2. 云数据处理平台框架图

4.3. 结果显示

结果显示环节是整个系统框架的最后阶段，具体又可以分为两个部分：数据可视化发展和数据分析应用。数据可视化是指图书馆对数据进行分析整合以及处理后，所得到的处理结果，不但可以被内部工作人员所查看，这可以内化为图书馆管理系统组成部分，而且可将可视化的原始数据提交给读者用户，直接为用户服务。数据分析应用既包括对数据本身的分析，也包括对内容节点联系的评估和对数据发展的预测。除此之外，数据分析的结果还可以为大数据环境下的图书馆的改进提供建议。

5. 大数据技术在图书馆中的创新

目前大多数图书馆凭借着馆内的资源或者各自共享的资源为用户们提供服务，对网络上的信息资源涉猎较少。大数据在生活和科研中越来越重要，用户们希望通过一种统一的检索平台来搜索获取到他们想要得到的数据信息。大数据技术有着非常重要的作用，它在对于学科服务、咨询服务等方面对图书馆起影响很深。为了满足大部分用户的需求，有必要对大数据技术展开更加深入的研究，对复杂信息资源

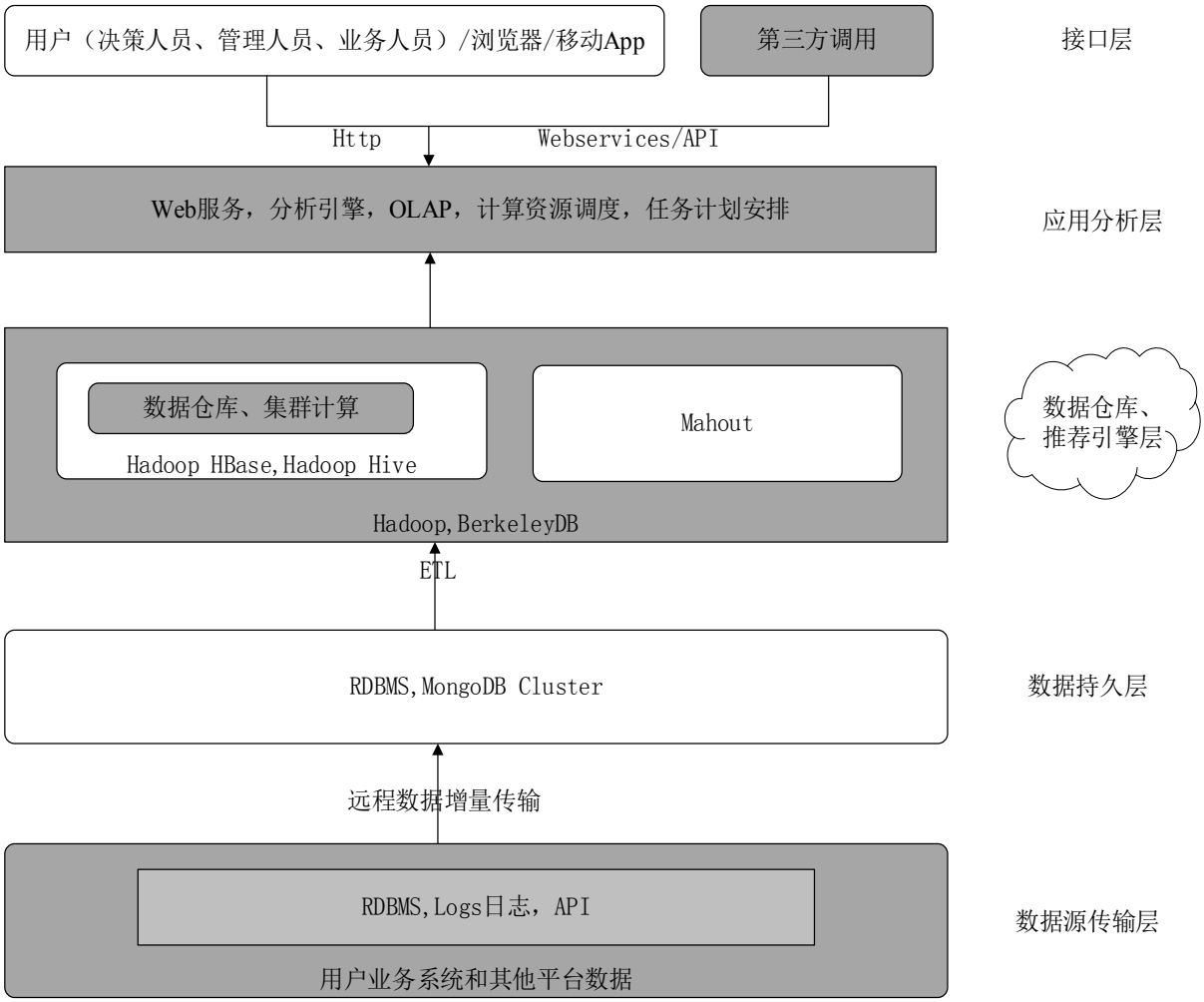

Figure 3. Hadoop system framework
图 3. Hadoop 系统架构

需要进行更加快速的分析处理。图书馆如何与数据库供应商以及其他用户进行数据共享和数据开发和维护，这将成为大数据环境下图书馆面临的一个重要问题。

众所周知，数字图书馆具有很多优点，如：信息资源数字化、信息利用共享化、信息提供知识化、信息传递网络化等。大数据环境下，数字图书馆结合各种先进的技术对数据进行处理，处理后的数据信息上传至数据库，通过数据库检索资源，以此来满足读者用户的信息需求。数据技术的发展与科学技术的不断进步，为图书馆的发展注入了新的活力。

6. 结论

图书馆作为技术敏感度极高的一类机构，管理部门应持续关注大数据的动态发展，积极思考图书馆发展过程中遇到的大数据问题，并尝试解决，这也是图书馆克服目前资源的局限，获得创新发展的关键所在。随着当今社会信息化程度的不断提高，图书馆的资源共享必将成为一种趋势，数字图书馆的发展将迎来前所未有的机遇和挑战，当下数字图书馆的建设已经成为当今许多图书馆建设的新兴重点。未来，在大数据环境下，互联网的方向将从“网页相联”走向“数据相联”和“知识相联”，大数据技术的应用必将是未来图书馆服务创新的重要领域。

参考文献 (References)

[1] 陆康, 刘慧, 周欣. 高校图书馆数据分析机制研究[J]. 知识管理论坛, 2015(4): 32-37.

[2] 宁耀莉. 大数据思维下高校图书馆学科服务创新机制探究[J]. 图书馆界, 2017(1): 1-4.

[3] 沈洋, 李小平. 大数据背景下高校图书馆学科服务的创新发展研究[J]. 新世纪图书馆, 2017(1): 46-49.

[4] 李杨, 韩洁茹, 等. “互联网+”时代高校图书馆学科服务策略研究[J]. 中国中医药图书情报杂志, 2016, 40(2): 10-13.

[5] 王晓宇. 大数据环境下高校图书馆学科服务创新初探[J]. 商, 2015(30): 297.

[6] 刘江红, 赵桂荣. 大数据时代高校图书馆个性化学科服务创新探究——以黑龙江大学图书馆为例[J]. 农业图书情报学刊, 2016, 28(12): 166-170.