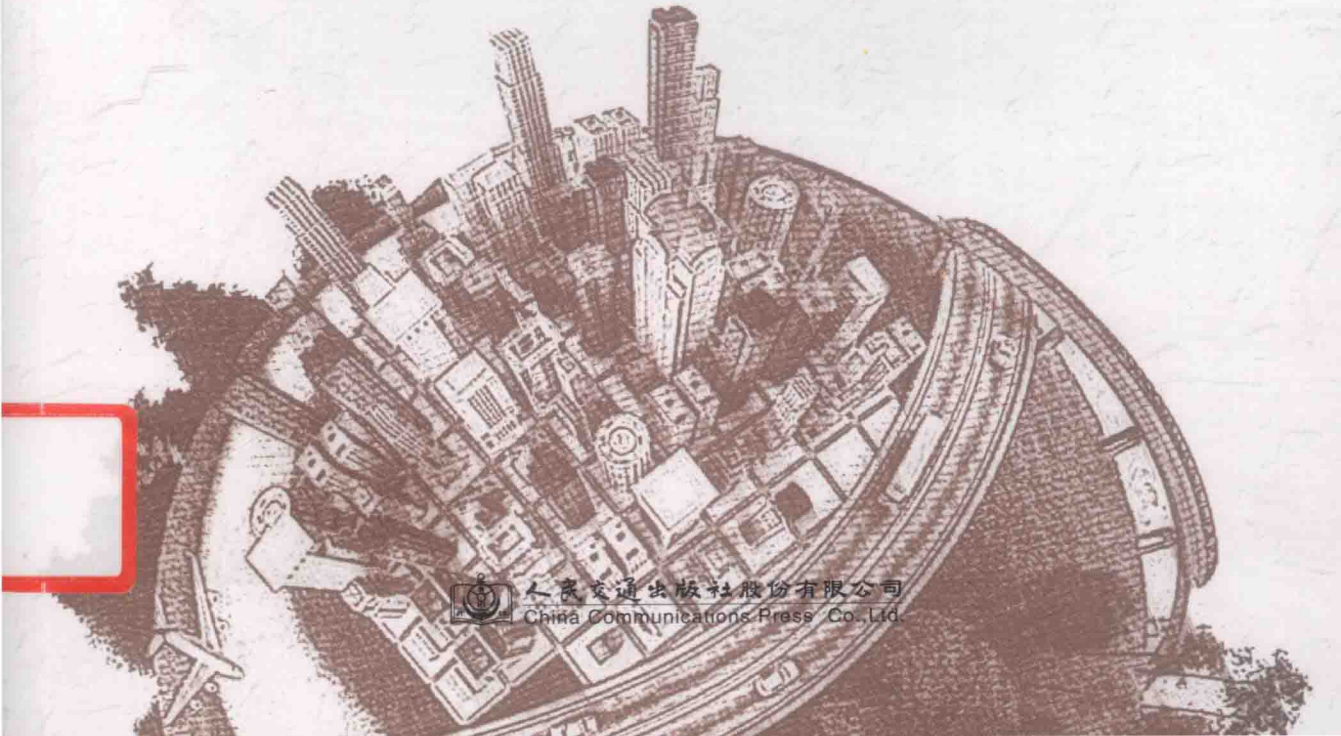
贵州省交通建设系列科技专著

交通大数据应用与实践

THE APPLICATION AND PRACTICE OF
TRANSPORTATION BIG DATA

贵州省交通运输厅 组织编写

罗 强 杨建国 康厚荣 程 洁 丁志勇 著



人民交通出版社股份有限公司
China Communications Press Co., Ltd.

贵州省交通建设系列科技专著

交通大数据应用与实践

贵州省交通运输厅 组织编写
罗 强 杨建国 康厚荣
程 洁 丁志勇 著

人民交通出版社股份有限公司
China Communications Press Co.,Ltd.

内 容 提 要

本书为“贵州省交通建设系列科技专著”中的一本。全书立足交通运输行业信息化应用发展实际，以打造“互联网＋交通”的智慧交通体系为愿景，深入浅出、循序渐进提出交通大数据发展背景、应用现状、建设需求、总体框架、关键技术和典型应用，系统阐述了大数据理念、技术、模式和应用。全书共6章，主要内容包括大数据发展概述、交通大数据采集体系、交通大数据中心、交通大数据处理平台、交通大数据分析应用和交通大数据建设运营模式。

本书适用于交通运输信息化规划、咨询、设计及建设管理人员，亦可供其他相关行业人员参考。

图书在版编目(CIP)数据

交通大数据应用与实践 / 罗强等著 ；贵州省交通运输厅组织编写. — 北京 ：人民交通出版社股份有限公司，2015.11

（贵州省交通建设系列科技专著）

ISBN 978-7-114-12580-5

Ⅰ. ①交… Ⅱ. ①罗… ②贵… Ⅲ. ①交通工程—数据处理 Ⅳ. ①U491

中国版本图书馆CIP数据核字(2015)第255407号

贵州省交通建设系列科技专著

书　　名：交通大数据应用与实践

著 作 者：罗　强　杨建国　康厚荣　程　洁　丁志勇

责任编辑：周　宇　牛家鸣

出版发行：人民交通出版社股份有限公司

地　　址：(100011)北京市朝阳区安定门外外馆斜街3号

网　　址：http://www.ccpress.com.cn

销售电话：(010)59757973

总 经 销：人民交通出版社股份有限公司发行部

经　　销：各地新华书店

印　　刷：北京市密东印刷有限公司

开　　本：787×1092　1/16

印　　张：9.75

字　　数：220千

版　　次：2015年11月　第1版

印　　次：2015年11月　第1次印刷

2016年2月　第2次印刷

书　　号：ISBN 978-7-114-12580-5

定　　价：45.00元

贵州省交通建设系列科技专著

编审委员会

总 序

Preface

古往今来，独特的地形地貌赋予贵州重峦叠嶂山高谷深的隽秀之美，但山阻水隔也桎梏着贵州经济社会发展的步伐。打破交通运输瓶颈，建设内捷外畅的现代综合交通运输体系，与全国同步迈向小康，一直是贵州人的夙愿。

改革开放特别是进入“十二五”以来，党中央、国务院及交通运输部等国家部委高度重视贵州经济社会发展。2012年年初，国务院出台支持贵州发展的国发2号文件，将贵州省经济社会发展的战略规划上升到国家层面。贵州省委、省政府立足当前、着眼长远，提出坚持把交通作为优先发展的重大战略，举全省之力加快交通基础设施建设。2012年以来，贵州省先后启动了高速公路建设、水运建设三年会战，普通国省干线公路建设攻坚，“四在农家·美丽乡村”小康路行动计划，“多彩贵州·最美高速”和“多彩贵州·平安高速”创建等一系列行动，志在“十二五”末，通过交通大建设一举打破大山的束缚，畅通经济发展的交通网络。

广大交通建设者紧紧抓住发展的历史机遇，凝心聚智，在广袤的黔山秀水之间，用光阴和汗水构筑贵州面向未来的交通新格局。“十二五”期间，全省交通基础设施建设将完成投资4 500亿元，新建成高速公路3 600公里，高速公路通车总里程将突破5 100公里，全省88个县（市、区）将全部通高速公路。乌江、赤水河建成四级航道700公里，改写了贵州无高等级航道的历史。建成构皮滩水电站翻坝枢纽工程，实现乌江航道全线通航。曾经的黔道天堑正变成康庄大道，一张以高速公路为骨架、国省干线公路为支撑、县乡公路为脉络、小康路为基础的四级公路路网正在形成，“扬帆赴江海”指日可待。

围绕贵州交通发展中出现的科技需求，贵州省交通运输厅组织开展了一批省部级重大科研项目攻关，重点突破一批关键、共性技术难题，在支撑工程建设、引领行业创新发展方面成效显著。在山区复杂条件下大型桥梁建设技术方面，形成了千米级悬索桥、高墩大跨刚构桥和钢管混凝土拱桥等设计施工成套技术，有力支撑了坝陵河大桥、清水河大桥、鸭池河大桥、赫章大桥、木蓬大桥等一批世界级桥梁建设工程，实现了我省桥梁建设技术的大跨越；针对西部山区复杂地质地形条件，从勘察设计、建设施工、养护管理和生态环保等方面系统开展基础研究和

技术开发，形成一批山区高速公路修筑技术，其成果居国内先进水平，有力支撑了复杂山区环境下高速公路项目建设；在山区航道整治、船型标准、通航枢纽建设等方面取得的创新性成果，促进了贵州航运工程的发展；完成了“贵州乌蒙山区毕都高速公路安全保障科技示范工程”等交通运输部科技示范项目，有力推动了交通科技成果推广应用；以“互联网＋便捷交通”推进智慧交通建设，率先开展智能交通云的建设和应用。交通运输科技成果连续3年获得贵州省科技进步和成果推广一等奖。

为展现在公路、水路和交通安全、信息化建设等方面取得的技术成就，促进技术交流，加大推广应用，贵州省交通运输厅组织编写了“贵州省交通建设系列科技专著”。这套科技专著的出版，对传承科技创新文化，提升交通科技水平，深入实施科技兴省战略，促进贵州经济社会快速发展，意义重大、影响深远。

交通成就千秋梦，东西南北贯黔中。编撰这套系列科技专著，付出的是艰辛、凝结的是智慧、反映的是成绩，折射了交通改变地理劣势、奋斗推动跨越的创新精神，存史价值较高，是一笔当代贵州的可贵财富。

2015年10月

前 言

Foreword

随着互联网、物联网、云计算等新一代信息技术的应用和推广，人类的衣食住行等时刻产生着数据，互联网数据每年以50%的速度增长，而目前世界上90%的数据都产生于近两年，我们迎来了大数据时代，人们开始关注大数据的应用研究。然而，身处大数据时代的我们，并没有做好迎接大数据的准备，从对大数据的模糊定义到对大数据应用价值的众说纷纭，都表明了大数据价值的实现与真正“落地”与我们的期望相去甚远。

交通大数据作为现代交通、智慧交通发展的产物，其战略意义不在于涵盖多少数据信息，而在于人们对它进行的更准确的挖掘和预测。如果把交通大数据当作一种产业，想要实现这种产业的盈利，就要把大量的交通数据通过“加工”转化成为有组织的信息、甚至是知识，实现数据的“增值”，诸如路况预测、风险规避、事故鉴定等。

全书围绕交通大数据应用与实践主题，以提升交通运输大数据治理能力为目标导向，从大数据采集、大数据聚集、大数据处理、大数据应用、大数据服务等多个维度，系统阐述了大数据技术在交通运输行业的应用发展情况，重点回答了大数据是什么、大数据框架怎么做、大数据如何应用、大数据推广服务等几个关键问题，分享了贵州、北京等地在交通大数据方面的典型实践和案例。

本书立足深入推进“智慧交通”建设，紧密围绕交通运输行业转型升级对交通大数据应用的迫切需求，以提升服务水平与治理能力为核心，以行业信息资源全面梳理、行业大数据分析能力建设为基础，以行业公共信息资源开放共享为理念，提出交通大数据技术框架和大数据若干关键技术，紧密结合交通运输行业数据资源特征特点，给出典型的交通行业大数据应用实践案例，引导社会优质力量参与行业大数据应用，服务公众、服务政府、服务企业，发挥交通大数据综合应用效益，激发行业大数据创新活力，为政府、行业信息化管理和应用领域的人员提供了一本框架性和系统性的大数据技术和应用实践指南。

全书分为六章。第一章为大数据发展概述，主要阐述大数据的发展背景、关键技术，介绍交通信息化发展概况和交通大数据发展面临的挑战；第二章为交通大数据采集体系，主要阐述物联网技术在智慧交通中的建设应用情况，给出贵州省智慧交通物联网的典型应用；第三章为

交通大数据中心，在研究国内其他行业数据中心建设应用状况的基础上，提出数据中心总体框架，并从数据采集处理、数据存储汇聚、数据交换共享、数据管控治理等多个维度描述交通大数据中心建设应用关键技术；第四章为交通大数据处理平台，主要结合贵州省智慧交通云建设应用情况，重点阐述了交通云计算基础设施、公共应用支撑平台、云应用服务等关键技术和应用实践；第五章为交通大数据分析应用，主要阐述大数据分析应用的核心价值、发展阶段、保障措施和存在问题，给出了大数据技术在智慧交通中的实际应用案例；第六章为交通大数据建设运营模式，主要研究了大数据建设运营模式、风险管理以及PPP模式的法律体系，结合贵州省交通大数据建设运营实际情况，归纳总结了典型大数据商业模式设计，为行业创新大数据应用发展模式提供一定借鉴。

本书由康厚荣主持编写，第一章由蹇峰、陈绍辉、韩悦完成；第二章由罗强、杨赞、程洁完成；第三章由杨建国、肖榕、程洁、郭锐、杨倩完成；第四章由康厚荣、丁志勇、程洁、金流鹏完成；第五章由程洁、李轶舜、魏攀一完成；第六章由李轶舜、程洁完成。

交通大数据应用是一个新生产业，由于作者理论水平和实践经验有限，书中内容难免有所欠缺，恳请广大读者批评指正。

作　者

2015年9月

目 录

Contents

第1章 大数据发展概述

1.1 大数据发展背景

近年来，包括大数据(Big Data)在内的信息发展极其迅速，不仅为智慧交通提供了技术基础，也对智慧交通的应用模式和发展理念产生了巨大影响。交通大数据已经成为世界各国的研究机构、商业企业及政府部门的关注热点，大量的研究和相关工作已经展开。大数据将在交通运行管理、出行信息服务、交通运输部门决策、交通应急和安全保障等领域发挥重大作用。

大数据的应用和技术是在互联网快速发展中诞生的，起点可追溯到2000年前后。当时互联网网页爆发式增长，每天新增约700万个网页，到2000年年底全球网页数达到40亿，用户检索信息越来越不方便。谷歌等公司率先建立了覆盖数十亿网页的索引库，开始提供较为精确的搜索服务，大大提升了人们使用互联网的效率，这是大数据应用的起点。当时搜索引擎要存储和处理的数据，不仅数量之大前所未有，而且以非结构化数据为主，传统技术无法应对。为此，谷歌提出了一套以分布式为特征的全新技术体系，即后来陆续公开的分布式文件系统(GFS，Google File System)、分布式并行计算(Map Reduce)和分布式数据库(Big Table)等技术，以较低的成本实现了之前技术无法达到的规模。这些技术奠定了当前大数据技术的基础，可以认为是大数据技术的源头。

伴随着互联网产业的崛起，这种创新的海量数据处理技术在电子商务、定向广告、智能推荐、社交网络等方面得到应用，取得巨大的商业成功。这一现象启发全社会开始重新审视数据的巨大价值，于是金融、电信等拥有大量数据的行业开始尝试这种新的理念和技术，取得初步成效。与此同时，业界也在不断对谷歌提出的技术体系进行扩展，使之能在更多的场景下使用。2011年，麦肯锡、世界经济论坛等知名机构对这种数据驱动的创新进行了研究总结，随即在全世界掀起了一股大数据热潮。

虽然大数据已经成为全社会热议的话题，但到目前为止，“大数据”尚无公认的统一定义。一般来说，可以从总体架构、共享交换、服务体系、领域应用等不同的角度和层面去认识和利用大数据。大数据是具有体量大、结构多样、时效强等特征的数据；处理大数据需采用新型计算架构和智能算法等新技术；大数据的应用强调以新的理念应用于辅助决策、发现新的知识，更强调在线闭环的业务流程优化。国际研究机构Gartner针对大数据给出了这样的定义：需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

从技术上看，大数据与云计算的关系就像一枚硬币的正反面一样密不可分。大数据必然无法用单台的计算机进行处理，必须采用分布式架构。它的特色在于对海量数据进行分布式数据挖掘，但它必须依托云计算的分布式处理、分布式数据库和云存储、虚拟化技术。应该说，随着云计算技术在交通运输领域的广泛应用，大数据将在智慧交通中发挥越来越重要的作用。

1.2 大数据技术架构

大数据来源于互联网、企业系统和物联网等信息系统，经过大数据处理系统的分析挖掘，产生新的知识用以支撑决策或业务的自动智能化运转。从数据在智慧交通中的生命周期看，大数据从数据源经过分析挖掘到最终获得价值一般需要经过 5 个主要环节，包括数据准备、数据存储与管理、计算处理、数据分析和知识展现，技术框架如图 1.1 所示。每个环节都面临不同程度的技术挑战。

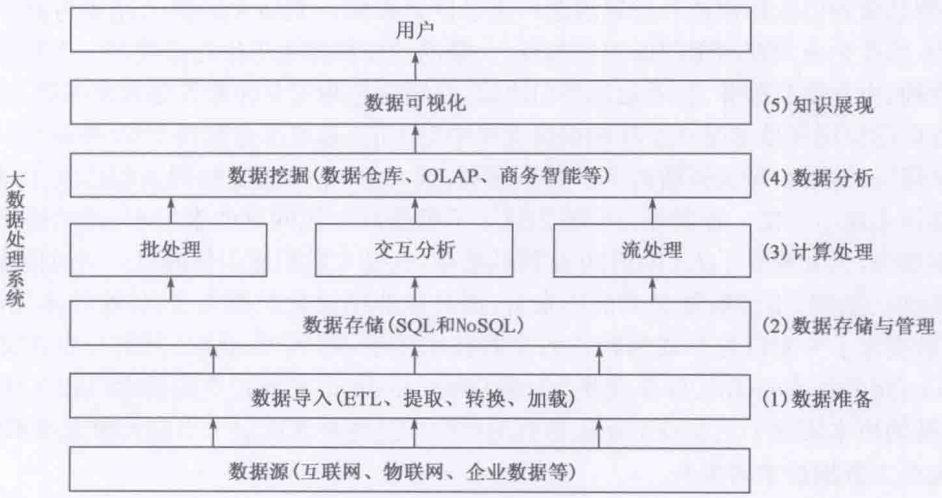


图 1.1 大数据技术框架

1.2.1 数据准备环节

在进行存储和处理之前，需要对数据进行清洗、整理，传统数据处理体系中称为 ETL(Extracting，Transforming，Loading)过程。与以往数据分析相比，大数据的来源多种多样，包括企业内部数据库、互联网数据和物联网数据，不仅数量庞大、格式不一，质量也良莠不齐。这就要求数据准备环节一方面要规范格式，便于后续存储管理，另一方面要在尽可能保留原有语义的情况下去粗取精、消除噪声。

1.2.2 数据存储与管理环节

当前全球数据量正以每年超过 50%的速度增长，存储技术的成本和性能面临非常大的压力。大数据存储系统不仅需要以极低的成本存储海量数据，还要适应多样化的非结构化数据

管理需求，具备数据格式上的可扩展性。

1.2.3　计算处理环节

需要根据处理的数据类型和分析目标，采用适当的算法模型，快速处理数据。海量数据处理要消耗大量的计算资源，对于传统单机或并行计算技术来说，速度、可扩展性和成本上都难以适应大数据计算分析的新需求。分而治之的分布式计算成为大数据的主流计算架构，但在一些特定场景下的实时性还需要大幅提升。

1.2.4　数据分析环节

数据分析环节需要从纷繁复杂的数据中发现规律、提取新的知识，这是大数据价值挖掘的关键。传统数据挖掘对象多是结构化、单一对象的小数据集，挖掘时更侧重根据先验知识预先人工建立模型，然后依据既定模型进行分析。对于非结构化、多源异构的大数据集的分析，往往缺乏先验知识，很难建立显式的数学模型，这就需要发展更加智能的数据挖掘技术。

1.2.5　知识展现环节

在大数据服务于决策支撑的场景下，以直观的方式将分析结果呈现给用户，是大数据分析的重要环节。如何让复杂的分析结果易于理解是该环节面临的主要挑战。在嵌入多业务中的闭环大数据应用中，一般是由机器根据算法直接应用分析结果而无须人工干预，这种场景下知识展现环节则不是必需的。

1.3　大数据关键技术

1.3.1　大数据存储管理技术

数据的海量化和快增长特征是大数据对存储技术提出的首要挑战。这要求底层硬件架构和文件系统在性价比上要大大高于传统技术，并能够弹性扩展存储容量。但以往网络附着存储系统(NAS)和存储区域网络(SAN)等体系，存储和计算的物理设备分离，它们之间要通过网络接口连接，这就导致在进行数据密集型计算(Data Intensive Computing)时，I/O 容易成为瓶颈。同时，传统的单机文件系统(如 NTFS)和网络文件系统(如 NFS)要求一个文件系统的数据必须存储在一台物理机器上，且不提供数据冗余性，可扩展性、容错能力和并发读写能力难以满足大数据需求。

谷歌文件系统(GFS)和 Hadoop 的分布式文件系统 HDFS(Hadoop Distributed File System)奠定了大数据存储技术的基础。与传统系统相比，GFS/HDFS 将计算和存储节点在物理上结合在一起，从而避免在数据密集计算中易形成的 I/O 吞吐量的制约，同时这类分布式存储系统的文件系统也采用了分布式架构，能达到较高的并发访问能力。大数据存储架构的变化如图 1.2 所示。

当前随着应用范围的不断扩展，GFS 和 HDFS 也面临瓶颈。虽然 GFS 和 HDFS 在大文

件的追加(Append)写入和读取时能够获得很高的性能,但随机访问(Random Access)、海量小文件的频繁写入性能较低,因此其适用范围受限。业界当前和下一步的研究重点主要是在硬件上基于 SSD 等新型存储介质的存储体系架构,同时对现有分布式存储的文件系统进行改进,以提高随机访问、海量小文件存取等性能。

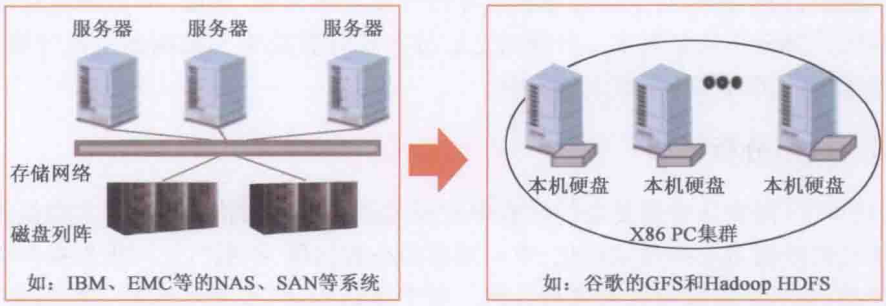


图 1.2　大数据存储架构的变化

大数据对存储技术提出的另一个挑战是多种数据格式的适应能力。格式多样化是大数据的主要特征之一,这就要求大数据存储管理系统能够适应对各种非结构化数据进行高效管理的需求。数据库的一致性(Consistency)、可用性(Availability)和分区容错性(Partition－Tolerance)不可能都达到最佳,在设计存储系统时,需要在 C、A、P 三者之间做出权衡。传统关系型数据库管理系统(RDBMS)以支持事务处理为主,采用了结构化数据表的管理方式,为满足强一致性(C)要求而牺牲了可用性(A)。

为大数据设计的新型数据管理技术,如谷歌 BigTable 和 Hadoop HBase 等非关系型数据库(NoSQL,Not only SQL),通过使用"键—值(Key-Value)"对、文件等非二维表的结构,具有很好的包容性,适应了非结构化数据多样化的特点。同时,这类 NoSQL 数据库主要面向分析型业务,一致性要求可以降低,只要保证最终一致性即可,给并发性能的提升让出了空间。谷歌公司在 2012 年披露的 Spanner 数据库,通过原子钟实现全局精确时钟同步,可在全球任意位置部署,系统规模可达到 100 万～1 000 万台机器。Spanner 不仅能够提供较强的一致性,而且支持 SQL 接口,代表了数据管理技术的新方向。整体来看,未来大数据的存储管理技术将进一步把关系型数据库的操作便捷性和非关系型数据库的灵活性结合起来,研发新的融合型存储管理技术。

1.3.2　大数据并行计算技术

大数据的分析挖掘是数据密集型计算,需要强大的计算能力。与传统"数据简单、算法复杂"的高性能计算不同,大数据的计算对计算单元和存储单元间的数据吞吐率要求极高,对性价比和扩展性的要求也非常高。传统依赖大型机和小型机的并行计算系统不仅成本高,数据吞吐量也难以满足大数据要求,同时靠提升单机 CPU 性能、增加内存、扩展磁盘等实现性能提升的纵向扩展(Scale Up)的方式也难以支撑平滑扩容。

谷歌在 2004 年公开的 MapReduce 分布式并行计算技术,是新型分布式计算技术的代表。一个 MapReduce 系统由廉价的通用服务器构成,通过添加服务器节点可线性扩展系统的总处

理能力(Scale Out),在成本和可扩展性上都有巨大的优势。谷歌的 MapReduce 是其内部网页索引、广告等核心系统的基础。之后出现的 Apache Hadoop 是谷歌 MapReduce 的开源实现,已经成为目前应用最广泛的大数据计算软件平台。MapReduce 架构能够满足"先存储后处理"的离线批量计算(Batch Processing)需求,但也存在局限性,最大的问题是时延过大,难以适用于机器学习迭代、流处理等实时计算任务,也不适合针对大规模图数据等特定数据结构的快速运算。

为此,业界在 MapReduce 基础上,提出了多种不同的并行计算技术路线。如 Yahoo 提出的 S4 系统、Twitter 的 Storm 系统是针对"边到达边计算"的实时流计算(Real Time Streaming Process)框架,可在一个时间窗口上对数据流进行在线实时分析,已经在实时广告、微博等系统中得到应用。谷歌 2010 年公布的 Dremel 系统,是一种交互分析(Interactive Analysis)引擎,几秒钟就可完成 PB(1PB=1 015B)级数据查询操作。此外,还出现了将 MapReduce 内存化以提高实时性的 Spark 框架、针对大规模图数据进行了优化的 Pregel 系统等。

针对不同计算场景建立和维护不同计算平台的做法,硬件资源难以复用,管理运维也很不方便,研发适合多种计算模型的通用架构成为业界的普遍诉求。为此,Apache Hadoop 社区在 2013 年 10 月发布的 Hadoop 2.0 中推出了新一代的 MapReduce 架构。新架构的主要变化是将旧版本 MapReduce 中的资源管理和任务调度功能分离,形成一层与任务无关的资源管理层(YARN)。如图 1.3 所示,YARN 对下负责物理资源的统一管理,对上可支持批处理、流处理、图计算等不同模型,为统一大数据平台的建立提供了新平台。基于新的统一资源管理层开发适应特定应用的计算模型,仍将是未来大数据计算技术发展的重点。

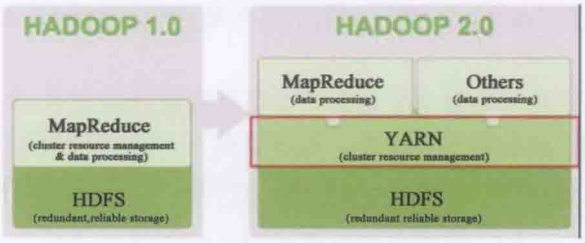


图 1.3　资源管理和任务调度功能分离的新式平台

1.3.3　大数据查询和分析技术

在人类全部数字化数据中,仅有非常小的一部分(约占总数据量的 1%)数值型数据得到了深入分析和挖掘(如回归、分类、聚类),大型互联网企业对网页索引、社交数据等半结构化数据进行了浅层分析(如排序)。对占总量近 60%的语音、图片、视频等非结构化数据还难以进行有效的分析。

大数据分析技术的发展需要在两个方面取得突破,一是对体量庞大的结构化和半结构化数据进行高效率的深度分析,挖掘隐性知识,如从自然语言构成的文本网页中理解和识别语义、情感、意图等;二是对非结构化数据进行分析,将海量复杂多源的语音、图像和视频数据转化为机器可识别的、具有明确语义的信息,进而从中提取有用的知识。

目前的大数据分析主要有两条技术路线,一是凭借先验知识人工建立数学模型来分析数据;二是通过建立人工智能系统,使用大量样本数据进行训练,让机器代替人工获得从数据中

提取知识的能力。由于占大数据主要部分的非结构化数据，往往模式不明且多变，因此难以靠人工建立数学模型去挖掘深藏其中的知识。

通过人工智能和机器学习技术分析大数据，被业界认为具有很好的前景。2006 年谷歌等公司的科学家根据人脑认知过程的分层特性，提出增加人工神经网络层数和神经元节点数量，加大机器学习的规模，构建深度神经网络，可提高训练效果，并且这一观点在后续试验中得到证实。这一事件引起工业界和学术界高度关注，使得神经网络技术重新成为数据分析技术的热点。目前，基于深度神经网络的机器学习技术已经在语音识别和图像识别方面取得了很好的效果。但未来深度学习要在大数据分析上广泛应用，还有大量理论和工程问题需要解决，主要包括模型的迁移适应能力，以及超大规模神经网络的工程实现等。

1.3.4 大数据可视化技术

数据可视化旨在借助于图形化手段，清晰有效地传达与沟通信息。但是，这并不意味着数据可视化就一定因为要实现其功能用途而令人感到枯燥乏味，或者是为了看上去绚丽多彩而显得极端复杂。为了有效地传达思想观念，美学形式与功能需要齐头并进，通过直观地传达关键的方面与特征，实现对于相当稀疏而又复杂的数据集的深入洞察。然而，设计人员往往并不能很好地把握设计与功能之间的平衡，从而创造出华而不实的数据可视化形式，无法达到其主要目的，也就是传达与沟通信息。

数据可视化的成功应归于其背后基本思想的完备性：依据数据及其内在模式和关系，利用计算机生成的图像来获得深入的认识和知识。其第二个前提就是利用人类感觉系统的广阔带宽来操纵和解释错综复杂的过程、涉及不同学科领域的数据集以及来源多样的大型抽象数据集合的模拟。

大规模数据的可视化和绘制主要是基于并行算法设计的技术，合理利用有限的计算资源，高效地处理和分析特定的数据集的特性。很多情况下，大规模数据可视化的技术通常会结合多分辨率表示等方法，以获得足够的互动性能。在面向大规模数据的并行可视化工作中，主要涉及 4 种基本技术：

(1)数据流线化(Data Streaming)：将大数据分为相互独立的子块后依次处理。在数据规模远大于计算资源时是一类主要的可视化手段。它能够处理任意大规模的数据，同时也能提供更有效的缓存使用效率，并减少内存交换，但通常这类方法需要较长的处理时间，不能提供对数据的交互挖掘。离核渲染(Out-of-Core Rendering)是数据流线化的一种重要形式。

(2)任务并行化(Task Parallelism)：把多个独立的任务模块平行处理。这类方法要求将一个算法分解为多个独立的子任务，并需要相应的多重计算资源。其并行程度主要受限于算法的可分解粒度以及计算资源中节点的数目。

(3)管道并行化(Pipeline Parallelism)：同时处理各自面向不同数据子块的多个独立的任务模块。对于任务并行化和管道并行化两类方法，如何达到负载的平衡是关键点。

(4)数据并行化(Data Parallelism)：将数据分块后进行平行处理，通常称为单程序多数据流(SPMD)模式。这类方法能达到高度的平行化，并且在计算节点增加时可以达到较好的可扩展性。对于非常大规模的并行可视化，节点之间的通信往往是制约因素。提高数据的本地性也可以大大提高效率。

以上这些技术往往在实践中相互结合，从而构建出一个更高效的解决方法。

可视化技术中图形的绘制是一个计算密集型的处理工作。在处理大规模数据时，使用可视化算法，以互动的速度来绘制图形已经超出了单一的 CPU 和 GPU 图形加速器的计算能力。数据并行绘制方法被普遍地用于提供可视化系统的交互速度。应用最普遍的并行绘制算法的分类是基于绘制流水线中图元排序的位置。

近年来受到关注的一种针对模拟计算产生的超大规模数据的可视化模式被称为原位可视化(In Situ Visualization)。它通过将模拟计算和可视化紧密结合，降低数据传输和存储成本。通常的可视化模式将 PB 量级模拟产生的全部数据传递到存储设备，再经处理后用于可视化。数据传输是整个系统的瓶颈，I/O 将占据绝大部分的计算时间。而在原位可视化模式中，数据之间在计算后原位被缩减与前处理，再用于随后的可视化与分析。经过缩减后的数据，通常比原始数据小多个数量级，能够极大地降低数据传输和存储的开支。

1.4　交通信息化发展概况

自 1975 年成立交通部(现为交通运输部，后同)计算机应用研究所至今，交通信息化的发展建设已历时 40 年。交通运输行业为适应交通运输发展需要，对交通运输信息化发展进行了不断探索，交通运输信息化经历了从无到有、从有到精、由点及面的发展历程，初步建成了日趋成熟的交通运输信息化体系，可分为三个不同特征的发展阶段。

1.4.1　单机应用阶段

1975 年，交通部计算机应用研究所成立，1989 年，出台了《交通运输经济信息系统(TEIS)——“八五”发展计划》，交通运输信息化开始起步，信息化理念得到普及。1989 年，交通部计算机应用研究所更名为中国交通信息中心，赋予其行业信息化管理职能，统筹交通行业的信息化建设和发展。20 世纪 70 年代，北京、上海、广州等大城市开始了交通信号控制的研究与开发，单点定周期交通信号控制器和线调信号控制系统进入公众视线。20 世纪 80 年代后期，我国开始尝试 ITS 基础性研究工作，尝试在交通信息采集、驾驶员考试培训、车辆动态识别等领域应用信息技术。

1.4.2　部门应用阶段

20 世纪 90 年代初，我国开始关注国际上 ITS 的发展，“九五”期间，交通部制定《公路、水运交通运输信息化“九五”规划和 2010 年远景目标》，开展交通运输信息网络(CTInet)建设；智能交通系统(ITS)理念被正式引入我国，ITS 框架标准和体系结构开始进入研究开发阶段。“十五”期间，无论公路、水路、运输，还是城市交通均得到了有力的信息化建设支持。《公路水路交通信息化“十五”发展规划》，提出围绕政府办公、行业监管、现代物流三大领域开展信息化建设。这段时期，交通运输行业基础信息网络基本完善，各部门业务应用逐步覆盖。2000 年，科技部会同国家计委、经贸委、公安部、铁道部、交通部、建设部、信息产业部等几十个部、委、局联合建立了“全国智能运输系统协调领导小组”及办公室，并成立了 ITS 专家咨询委员会。