流式大数据实时处理技术、平台及应用

陈纯
浙江大学计算机科学与技术学院，浙江 杭州 310058

摘要
大数据处理系统根据其时效性可分为批式大数据和流式大数据两类。上述两类系统均无法满足“事中”感知查询分析处理模式的需求。为此，从分析大数据应用场景入手，提出了“流立方”流式大数据实时处理技术和平台，在完整大数据集上实现了低迟滞、高实时的即席查询分析。目前基于“流立方”平台开发的业务系统已应用到金融风控反欺诈、机器防御等领域，具有广阔的应用前景。

关键词
流式大数据；流处理；增量计算；时序处理

中图分类号：TP319　　文献标识码：A　　doi: 10.11959/j.issn.2096-0271.2017036

Real-time processing technology, platform and application of streaming big data

CHEN Chun
College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China

Abstract
According to its timeliness, big data processing systems can be categorized into two groups, namely batching big data processing and streaming big data processing. Both systems mentioned above are unable to meet the real-time requirement for censoring and query analysis tasks. To this end, the “stream cube” real-time data analysis technology and platform were presented, which can perform timely query with low lag. Currently, this technology has been applied to many fields, including financial risk management, anti-fraud as well as web bots defense, and offers promising prospects for further applications.

Key words
streaming big data, streaming processing, incremental computation, time series processing

1 引言

大数据技术的广泛应用使其成为引领众多行业技术进步、促进效益增长的关键支撑技术。根据数据处理的时效性，大数据处理系统可分为批式（batch）大数据和流式（streaming）大数据①两类。其中，批式大数据又被称为历史大数据，流式大数据又被称为实时大数据。

目前主流的大数据处理技术体系主要包括Hadoop[1]及其衍生系统。Hadoop技术体系实现并优化了MapReduce[2]框架。Hadoop技术体系主要由谷歌、推特、脸书等公司支持。自2006年首次发布以来，Hadoop技术体系已经从传统的“三驾马车”（HDFS[1]、MapReduce和HBase[3]）发展成为包括60多个相关组件的庞大生态系统②。在这一生态系统中，发展出了Tez、Spark Streaming[4]等用于处理流式数据的组件。其中，Spark Streaming是构建在Spark基础之上的流式大数据处理框架。与Tez相比，其具有吞吐量高、容错能力强等特点，同时支持多种数据输入源和输出格式。除了Spark开源流处理框架，目前应用较为广泛的流式大数据处理系统还有Storm[5]、Flink[6]等。这些开源的流处理框架已经被应用于部分时效性要求较高的领域，然而在面对各行各业实际而又差异化的需求时，这些开源技术存在着各自的瓶颈。

在互联网/移动互联网、物联网等应用场景中，个性化服务、用户体验提升、智能分析、事中决策等复杂的业务需求对大数据处理技术提出了更高的要求。为了满足这些需求，大数据处理系统必须在毫秒级甚至微秒级的时间内返回处理结果。以国内最大的银行卡收单机构银联商务为例，其日交易量近亿笔，需对旗下540多万个商户进行实时风险监控，在确保这些商户合规开展收单业务的同时，最大限度地保障个人用户的合法权益。这样的高并发、大数据、高实时应用需求给大数据处理系统提出了严峻的挑战。银联商务以前使用的$T+1$事后风控系统存在风险侦测迟滞高（次日才能发现风险，损害已经造成）、处理时间长（十几个小时之后才能完成风险识别）、无法处理长周期历史数据（只能分析最近几日的流水数据）以及无法支持复杂规则（仅能支持累积求和等简单规则）等重大缺陷。为此，亟须研发全新的事中风控系统，以重点实现低迟滞（在1 min内甄别突发风险）、高实时（100 ms内返回处理结果）、长周期（可处理长达10年以上的历史周期数据）以及支持高复杂度规则（如方差、标准差、K阶中心矩、最大连续统计等）等目标。这一目标可以抽象为一个大数据处理科学问题：如何在一个完整的大数据集上，实现低迟滞、高实时的即席（Ad-Hoc）查询分析处理。

2 技术解析

现有的大数据处理系统可以分为两类：批处理大数据系统与流处理大数据系统。以Hadoop为代表的批处理大数据系统需先将数据汇聚成批，经批量预处理后加载至分析型数据仓库中，以进行高性能实时查询。这类系统虽然可对完整大数据集实现高效的即席查询，但无法查询到最新的实时数据，存在数据迟滞高等问题。相较于批处理大数据系统，以Spark Streaming、Storm、Flink为代表的流处理大数据系统将实时数据通过流处理，逐条加载至高性能内存数据库中进行查询。此类系统可以对最新实时数据实现高效预

① https://www.infoq.com/articles/stream-processing-hadoop/

② http://dbaplus.cn/news-21-288-1.html

设分析处理模型的查询，数据迟滞低。然而受限于内存容量，系统需丢弃原始历史数据，无法在完整大数据集上支持Ad-Hoc查询分析处理。因此，研发具有快速、高效、智能且自主可控特点的流式大数据实时处理技术与平台是当务之急。

实现一个融合批处理和流处理两类系统且对应用透明的系统级方案，需要攻克以下几个技术难点。

（1）复杂指标的增量计算

尽管计数、求和、平均等指标能够依靠查询结果合并实现，然而方差、标准差、熵等大部分复杂指标无法依靠简单合并完成查询结果的融合。再者，当查询涉及热点数据维度及长周期时间窗口的复杂指标时，多次重新计算会带来巨大的计算开销。

（2）基于分布式内存的并行计算

采用粗放的调度策略（例如约定在每天的固定时间将流数据导入批处理系统）会造成内存资源的极大浪费，亟须研究实现一种细粒度的基于进度实时感知的融合存储策略，以极大地优化和提升融合系统的内存使用效率。

（3）多尺度时间窗口漂移的动态数据处理

来自业务系统的数据查询请求会涉及多种尺度的时间窗口，如“最近5笔刷卡交易的金额”“最近10 min内密码重试次数”“过去10年的月均交易额”等。每次查询请求都重新计算结果会对系统性能造成极大的影响，亟须研究实现一种支持多种时间窗口尺度（数秒到数十年）、多种窗口漂移方式（数据驱动、系统时钟驱动）的动态数据实时处理方法，以快速响应来自业务系统的即席查询请求。

（4）高可用、高可扩展的内存计算

基于内存介质能够大大提升数据分析及处理能力，然而由于其易挥发的特性，一般需要采用多副本的方式来实现基于内存的高可用方案，这使得“如何确保不同副本的一致性”成为一个待解决的问题。此外，在集群内存不足或者部分节点失效时，“如何让集群在不间断提供服务的同时重新平衡”同样是一个待解决的技术难题。亟须研究分布式多副本一致性协议以及自平衡的智能分区算法，以进一步提升流处理集群的可用性以及可扩展性。

“流立方”流式大数据实时处理技术在上述领域取得了一系列突破，该技术提供基于时间窗口漂移的动态数据快速处理，支持计数、求和、平均、最大、最小、方差、标准差、K阶中心矩、递增/递减、最大连续递增/递减、唯一性判别、采集、过滤等多种分布式统计计算模型，并且实现了复杂事件、上下文处理等实时分析处理模型集的高效管理技术。

3 平台纵览

基于“流立方”流式大数据实时处理技术，研发了“流立方”流式大数据实时处理平台。其应用框架如**图1**所示，具有良好的灵活性和适应性。平台的数据装载模块负责从具体业务系统中接入实时流数据，数据抽取模块负责批量抽取历史数据，模型装载模块负责将分析处理模型集中的计算模型和脚本加载到平台中。当收到业务系统发出的实时查询请求时，“流立方”平台能够根据分析处理模型在完整大数据集上实时计算出相应的指标，并进行判断，将结果反馈给业务系统。

在测试环境为8台服务器（每台服务器配置24核 CPU、256 GB内存），同时计算16个统计指标（涉及4个维度，包含计数、求和、平衡、最大、最小、标准差、过滤、去重、排序、复杂事件处理等多种算法）的性

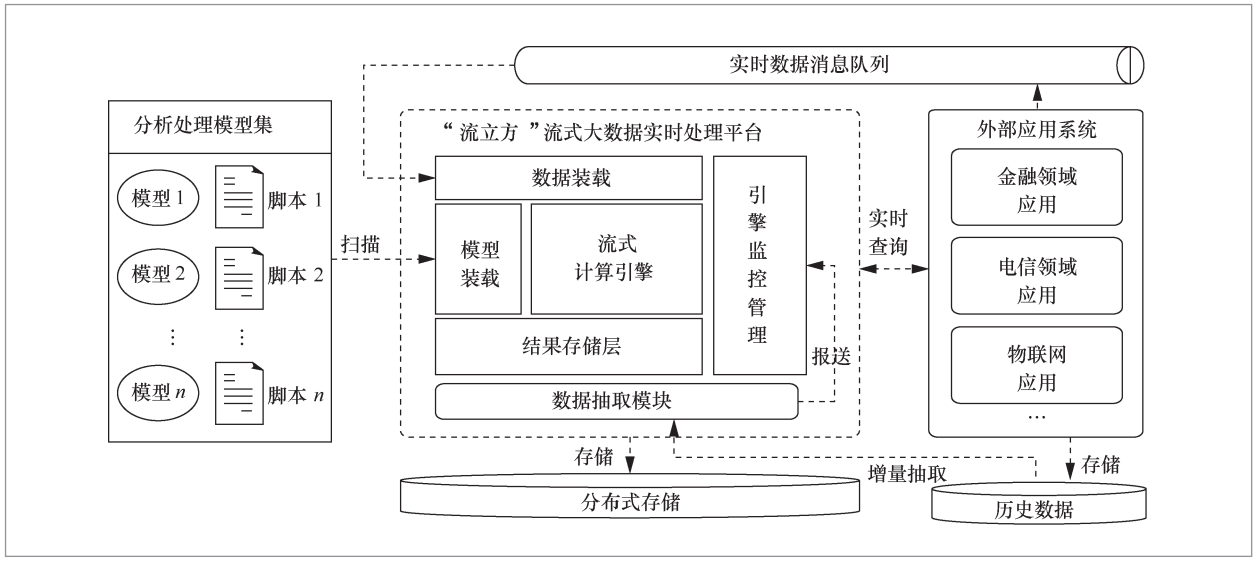

图 1 “流立方”平台应用框架

能测试中，“流立方”平台达到了单节点写入大于43 000 TPS、8节点读取大于100万TPS、平均时延为1~2 ms的优异性能，如图2所示。

“流立方”平台在解决批式大数据和流式大数据融合实时处理技术难题，实现优异性能的同时，还解决了流式大数据处理平台面临的两大工程化难题。一是作业的编排效率问题。大部分开源流处理平台在完成一个流处理编排时，都需要经过拓扑设计、代码编写、功能测试、打包部署等环节，一般需要一周的时间才能完成。“流立方”平台通过基于“所见即所得”的在线作业编排管理，将上线任务耗时降低到分钟级，大大提升了流处理作业的编排效率。二是流处理作业的灵活变更问题。流处理平台擅长进行逻辑预先定义的增量计算，尽管其计算效率极高，但计算灵活度受到限制。例如，某业务需要统计过去3个月的数据，现有的流处理平台在该业务上线3个月后才能完全生效，这样的工作方式使流处理技术在实际应用中受到很大的局限。“流立方”平台创新性地引入流媒体播放器的录制与重放思路，在原始数据进入流处理平台时，通过顺序写的方式持久化一份原始数据，在需要上线新的计算作业时，即刻重发指定时间窗口内的原始数据，从而实现快速（分钟级甚至秒级）计算作业上线。

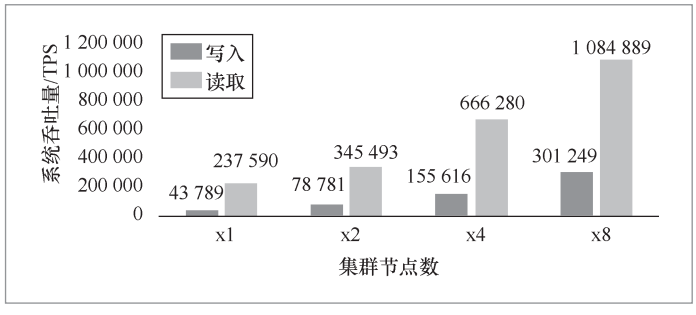

图 2 “流立方”平台性能指标

“流立方”平台引入了一系列创新技术，在性能、可用性、可扩展性等多个层面提升了流处理平台的处理能力，满足金融领域在内的众多领域的业务及运维需求。引入数据冲突智能规避技术，解决了流式处理中的热点数据处理问题，从而解决了大颗粒数据维度的处理效率问题；引入Paxos一致性协议，解决内存存储计算时

多副本一致性问题，提供了面向运维人员透明的一致性解决方案；引入智能分区技术，基于一致性散列技术，进一步将散列值拆解为散列块，通过散列块的平滑迁移解决存储集群的可伸缩性设计问题，确保对于运维人员的集群变更透明性；引入计算作业的动态运行时加载技术，规避了作业手工打包部署的问题，进一步提升了开发人员的工作效率。

在国内某大型银行卡收单机构组织的招标测试中，测试环节为两台低配置虚拟机，测试数据为该机构的数千万笔交易流水，计算逻辑包括50多条规则，涉及30多个统计指标。在该测试环节下，两家国外著名厂商中，一家厂商的计算时间长达24 h，另一家老牌数据库软件提供商则未能在一天内完成计算。相较于这些国外著名厂商的大数据处理平台，“流立方”平台能够在3 h内完成所有计算，且正确率为100%。

4 应用场景

“流立方”流式大数据实时处理系统在金融、交通、电信、公安等行业具有广泛的应用场景。以金融风控反欺诈为例，部署“流立方”风控系统仅需在交易前端增加风控探头，将实时交易数据旁路接入系统。“流立方”风控系统根据融合了专家知识和机器学习结果的数百条规则对每笔交易进行风险评估，判断是否允许进行该笔交易，流程如**图3**所示。该系统平均响应时间在6 ms以下，并发数超过50 000笔/s。同时，实现这一性能仅需要4台服务器。

基于“流立方”的金融风控反欺诈技术体系包含技术（如设备指纹、代理侦测、生物识别、关联分析、机器学习等技术）、知识（如盗卡反欺诈、伪卡反欺诈、信用卡

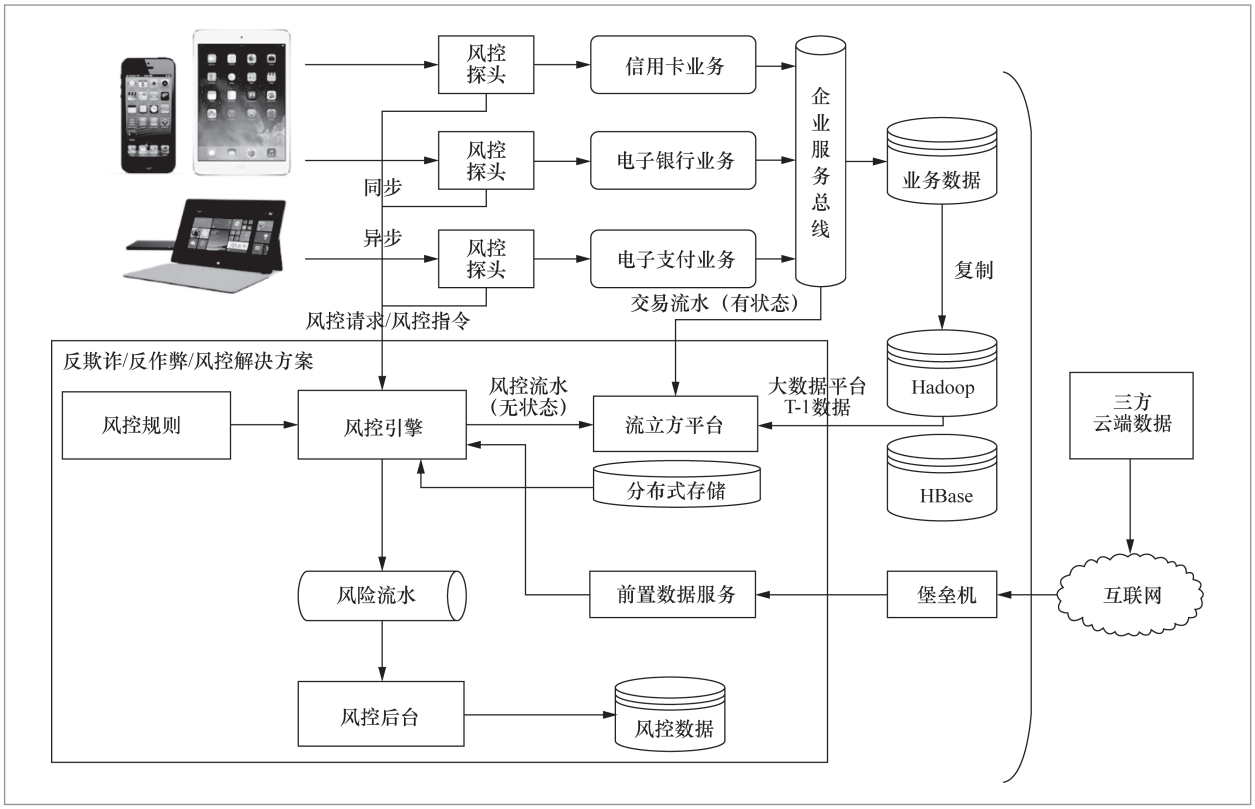

图 3 基于“流立方”的金融风控反欺诈流程

套现、营销反欺诈等规则与模型）、数据（如虚假手机数据、代理IP数据、P2P失信数据等标识数据）三大板块。技术部分中的设备指纹技术通过主被动混合的形式采集设备中软硬相关要素，结合概率论等算法为每一个设备颁发一个全球唯一的指纹编码，这些指纹编码在反欺诈的整个过程中起到非常积极的作用；代理侦测技术通过短时间内扫描IP相关端口来识别那些开启代理的IP，并在这些IP访问金融服务时进行识别；生物识别技术通过采集设备上用户的鼠标点击、触摸、键盘敲击等行为识别操作者是人还是机器以及是否操作者本人的问题；关联分析技术在底层通过图数据库存储不同节点以及关系信息，最终在界面上通过图的形式进行欺诈者关联分析及复杂网络分析；机器学习技术通过有监督、无监督的机器学习算法提升欺诈识别的准确率及覆盖率，并结合流立方技术提供模型的事中预测能力。

基于上述技术体系，研发了银行业务风险实时监控系统、互联网支付业务风险实时监控系统、电商业务风险实时监控系统等金融风控反欺诈系列解决方案。这些方案已应用到银行、第三方支付机构、互联网金融等领域的上百家企业。目前50%以上的线下交易都在“流立方”的保护下进行，基于“流立方”的金融风控反欺诈解决方案每天为我国的金融机构抵御上亿次的攻击。该技术已经成为我国金融安全领域基础设施必不可少的组成部分。

此外，在互联网机器防御系统中，“流立方”同样能发挥巨大作用。如今网络机器人遍布票务、电商、招聘、银行、政府、社交等各类网站，消耗了40%~60%的网络流量[③]。网络机器人不仅消耗网络资源、影响正常客户访问、增加网站运营成本，还会爬取产品、价格信息，形成不正当竞争，甚至混淆网站用户生态，影响营销分析。传统的控制策略通过采取屏蔽频繁访问、设置验证码等方式防御网络机器人，无法应对日益智能化的新型网络机器人。基于“流立方”的互联网机器防御系统通过在Web服务器上嵌入插件或者独立的嗅探器（sniffer）程序，将全流量的Web访问请求旁路到独立的机器防御集群，进行实时的流量分析及防御决策，并将决策后的结果实时回馈到Web服务器插件中。Web服务器插件在判定当前访问的设备或者IP地址等是机器人时，能够自动改写响应内容，根据不同的风险级别自动拒绝交易或将访问者引导到第三方图形验证码服务商进行机器人验证。访问者在通过验证后可以继续正常访问Web服务。该系统还创新地将设备指纹以及人机识别服务运用到机器防御系统中，不仅增加了可分析维度，提升了控制颗粒度，同时能够对基于浏览器内核的高级爬虫进行防护。此外，将机器防御规则、数据服务、设备指纹、人机识别以及图形验证码以软件即服务（software as a service，SaaS）的形式提供服务，进一步降低了互联网网站客户的运维门槛，提升了产品竞争力。该机器防御系统工作过程如**图4**所示。

基于“流立方”的实时机器防御系统通过多服务器访问流水关联决策、长周期数据决策、复杂规则爬虫识别、设备维度爬虫识别、人机识别等技术，实现了微秒级（400~800 μs）的识别时延，同时具有机器人识别管控一体化、轻量级接入等优点。根据已经接入机器防御服务的几十家客户的反馈，基于“流立方”平台的防御系统对机器人识别覆盖率在95%以上，准确率为99.9%。该机器防御系统能够拦截这些客户业务系统中占原有访问总流量80%~90%的来自网络机器人的访问流量，使得其业务系统服务器的压力降为原来的10%。由于基于“流立方”的机器防御系统

③ https://www.incapsula.com/blog/bot-traffic-report-2013.html

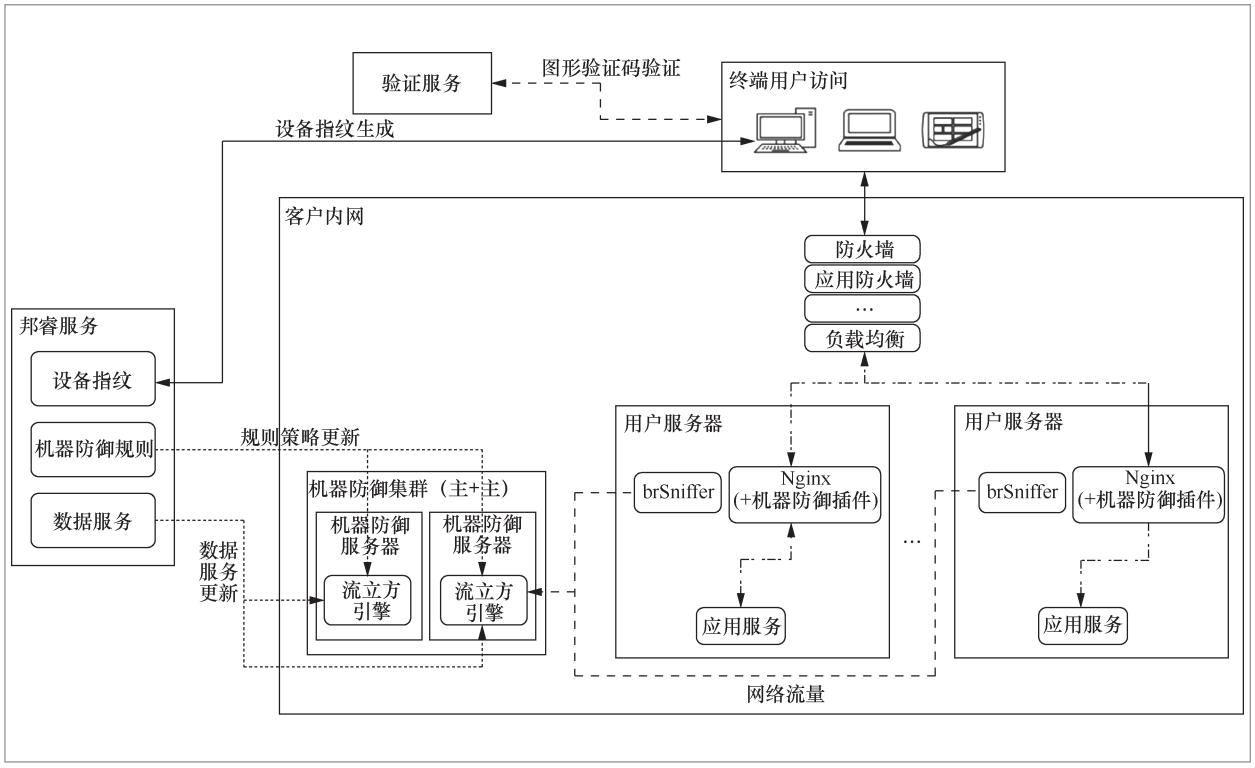

图 4 机器防御系统架构

的卓越识别及控制机器人的能力，当前，全国最大的票务平台正在对此服务进行全面的测试，希望能够进一步提升其票务服务能力。

此外，基于“流立方”的流式大数据实时处理平台在智慧交通领域也大有作为。通过实时分析从预埋在全国各地的摄像头采集的车牌信息，配合地理位置信息服务以及基于地理信息系统（geographic information system，GIS）的最短交通距离计算，实现实时套牌车信息抓取，为进一步打击违法犯罪服务提供帮助；通过实时分析交叉路口双向的车流量信息，实时控制每个路口的红绿灯、智能变换潮汐车道及可变车道，从而大大提升城市的通行效率。

“热数据”带来无与伦比的价值，数据从产生开始，其应用价值随时间的流逝呈现指数式下降，如何充分应用“热数据”是一个新生事务，是一个长期任务，也是流式大数据处理技术大有可为之处。“流立方”流式大数据实时处理技术和平台在金融、电信、交通、公安、海关、网络安全等需要引入“事中”感知分析决策模式的行业都具有广阔的应用前景。

5 结束语

基于批式大数据，可以不断学习新的知识，累积新的经验。然而，在应用这些知识和经验时，流式大数据更能够极大限度地挖掘“热数据”的潜在价值。这使得流式大数据技术具备更有效的应用推广价值。

流式大数据实时处理是大数据时代信息化的重要抓手。采用“事中”甚至“事前”模式实现感知、分析、判断、决策等功

能的智能系统需要流式大数据实时处理平台的支撑。此外，流式大数据实时处理可以为大数据驱动的深度学习提供计算框架支撑。“流立方”流式大数据实时处理平台可为研制融合逻辑推理、概率统计、众包、神经网络等多种形态的下一代人工智能统一计算框架提供支持。

参考文献:

[1] SHVACHKO K, KUANG H, RADIA S, et al. The Hadoop distributed file system[C]// 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), May 3-7, 2010, Incline Village, NV, USA, USA. New Jersey: IEEE Press, 2010: 1-10.

[2] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1): 107-113.

[3] CHANG F, DEAN J, GHEMAWAT S, et al. Bigtable: a distributed storage system for structured data[J]. ACM Transactions on Computer Systems (TOCS), 2008, 26(2): 4.

[4] ZAHARIA M, DAS T, LI H, et al. Discretized streams: an efficient and fault-tolerant model for stream processing on large clusters[J]. HotCloud, 2012(12): 10.

[5] TOSHNIWAL A, TANEJA S, SHUKLA A, et al. Storm@ twitter[C]// The 2014 ACM SIGMOD International Conference on Management of Data, June 22-27, 2014, Snowbird, Utah, USA. New York: ACM Press, 2014: 147-156.

[6] CARBONE P, KATSIFODIMOS A, EWEN S, et al. Apache flink: stream and batch processing in a single engine[J]. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2015, 36(4): 28-38.

作者简介

陈纯(1955-)，男，博士，浙江大学计算机科学与技术学院教授，中国工程院院士，计算机应用专家，主要研究方向为大数据智能计算、计算机图形图像处理等。

收稿日期：2017-05-16