

“双十一”流量大考过关 流计算功不可没

本报记者 崔爽

如何扛住流量“洪峰”，是每年“双十一”的技术大考。随着流量的飙升，这一大考无疑变得越来越难。

“双十一”刚过，各大电商纷纷公布当日战绩。从目前公布的数据不难看出，今年“双十一”的订单创建量峰值创下新高，单日数据处理量纪录也再次被刷新。

“双十一”的数据量不仅大，还会实时变化。可以说，每个购物数据都是个性化的、动态的，完成处理它们的工作，需要有强大的计算能力作为支撑。

“今年的数据处理工作，除了批处理（对数据进行批量处理）外，还有流处理，就是实时处理数据。庞大的交易数据每分钟、每秒钟都在变，它们并不是从数据

库里面统计出来的，而是系统自动一层层把它们汇集上来的。”阿里巴巴集团首席技术官张建锋表示。他口中的“流处理”就是流计算，即实时计算。对于扛住“双十一”流量“洪峰”，它功不可没。

那么，什么是流计算？它是如何扛住“双十一”流量“洪峰”的？对此，科技日报记者采访了业内专家。

以内存取代硬盘实现快速处理

北京理工大学计算机学院副院长、教授刘驰在接受科技日报记者采访时解释，流计算指当一条数据被处理后，立刻被序列化到内存中，然后通过网络传输到下一个节点，由下一个节点继续处理。“这种流式处理技术，是以内存取代硬盘的方式来实现数据快速处理，这是流计算之所以能够高效处理数据的根本原因。”他表示。

据刘驰介绍，流计算具有高性能、海量式、实时性、分布式、易用性、可靠性等

多项优势，主要的应用方向有，对金融与科学计算中的数据进行快速运算和分析；对存在于社交网站、电子邮件、视频、电话记录、电子感应器中的数字格式信息进行快速处理并反馈等。

“流式处理可被用于两种不同的计算场景：事件流和持续计算。”刘驰介绍道，在事件流场景，系统会持续产生大量的数据，这类数据最早出现在银行和股票交易领域，也会在互联网监控、无线通信等领域出现。这些领域的相关业务，

需要流计算技术以近乎实时的方式对数据流进行复杂分析。

“而在持续计算场景，比如大型网站，流计算技术可以动态实时地刷新用户访问数据，展示网站实时流量的变化情况，分析每天各小时的流量和用户分布情况。一些需要实时处理数据的场景也能应用到流计算技术，比如对根据用户行为产生的日志文件进行实时分析，对用户进行商品的实时推荐等。”刘驰介绍道。

实时采集、实时计算、实时查询

刘驰向记者介绍道，流计算的流程一般包含三个阶段：数据实时采集、数据实时计算和实时查询服务。

首先，后台工作人员会用实时数据集成工具，将数据实时传输到流式数据存储系统。具体过程是，系统将长时间累积的大量数据，平摊到每个计算时间节点，数据会被不停地进行小批量实时传输。此时，数据将会源源不断地写入流数据存储系统，不需要预先加载的过程。同时，在此过程中，数据是持续流动的，在计算完成

后就会被立刻丢弃。

然后，不同于批量计算等待数据集全部完成后才启动计算作业，流式计算作业是一种常驻计算服务，一旦启动将一直处于等待事件触发的状态，即一有小批量数据进入流式数据存储系统，流计算系统就会立刻计算并得出结果。“同时，部分电商平台的流计算团队，还使用了增量计算模型，将大批量数据分批进行增量计算，进一步减少单次运算规模并有效降低整体运算时延。”刘驰说。

最后，采用数据批处理方式，通常需要等待数据计算结果得出后，才能批量将数据传输到在线系统。而流式处理方式可在每次小批量数据计算工作完成后，就立刻将计算结果写入在线系统。“这样无需等待全部数据计算工作完成，我们就可在流计算数据查询系统中，查询到在线系统的数据处理结果。技术人员也可将实时结果发送给可视化系统，以实现计算结果的实时化展示。”刘驰说。

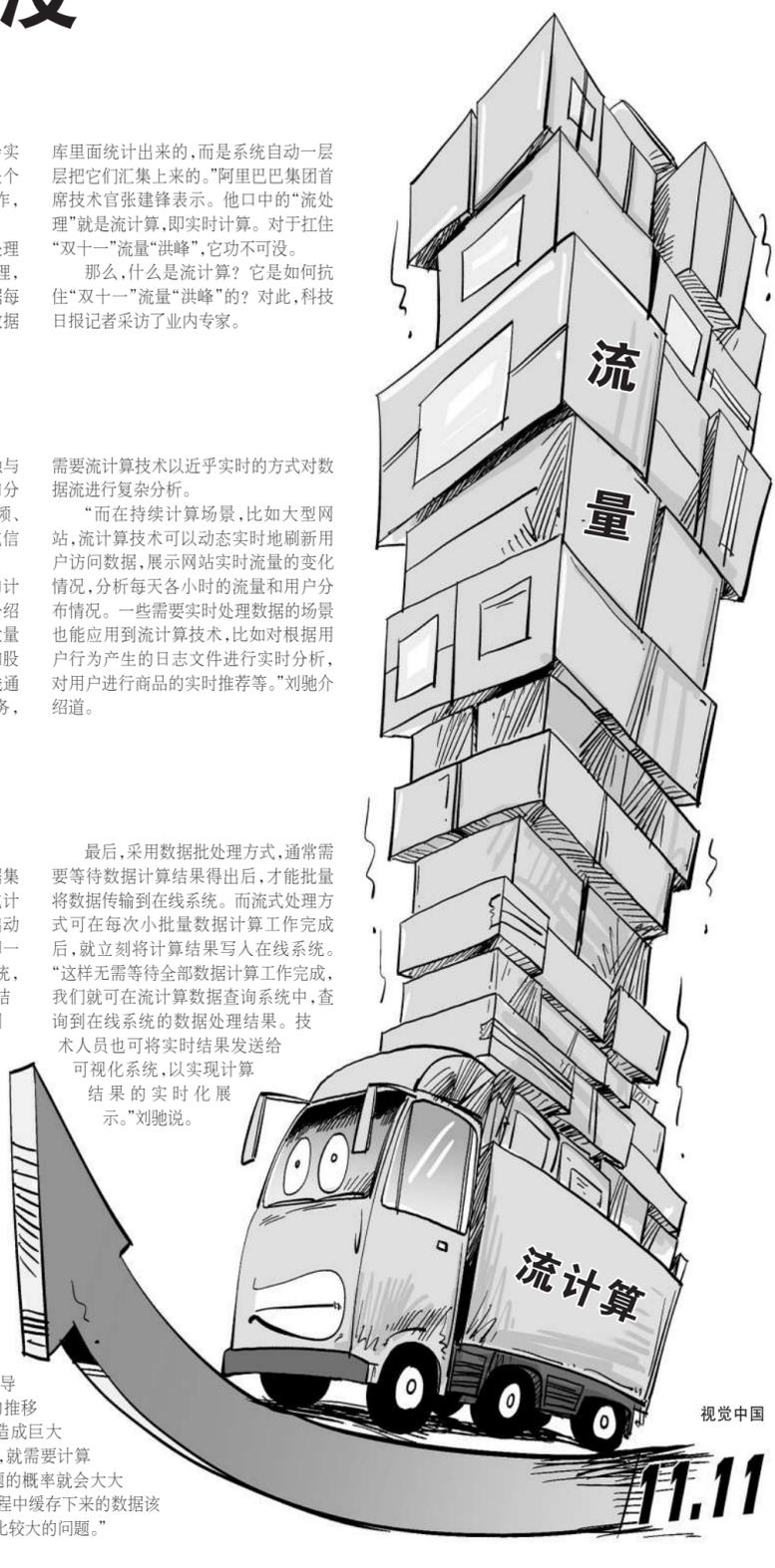
技术虽强大但仍存在短板

不过，刘驰表示，流计算虽然强大，也存在技术短板。

“如果在系统内存不足的情况下，还要满足多个用户的数据请求，那么每个用户实际被分到的内存就会很少。此时，应用流式处理技术后，由于内存需求量增加，系统会把原来分给多个用户的内存资源，全部分给一个用户。其他需要内存的用户，则会因为没有内存可用，被迫进入等待状态，直到那个占有内存资源用户退出后，其他用户才能使用内存。”刘驰说，同时，实时的计算环境会由

于各种各样的原因，比如网络延时等，导致程序到达计算节点顺序出现变化。

“此外，流计算技术面对的是源源不断涌入的数据，而如果还像批处理那样来计算数据的话，就可能会导致计算根本无法结束，或随着时间的推移需要保存的数据越来越多，给内存造成巨大压力。”刘驰补充道，“数据不断产生，就需要计算持续进行。但计算时间一长，出问题的概率就会大大增加。一旦出现问题，一些在计算过程中缓存下来的数据该怎么长期保存、怎么恢复就会成为比较大的问题。”



数据传输用上5G，存储不能停在4G

本报记者 张蕴

凭借超大带宽和超快的传输速度，5G将为海量数据的传输提供便利，为“大块头”数据文件提供“闪送”服务。

据媒体近日报道，有业内人士指出，目前5G的峰值传输速度或达1Gbps，未来随着毫米波技术的引入，这一速度有望超过10Gbps。相比速度的步步高升，目前的存储环境，却显得有些不匹配，相关设备恐难以“消化”闪送而来的海量数据。有人甚至戏言，数据传输用上了5G，但存储似乎还停在4G时代。

采集、存储、计算、分析，这是大型数据文件一般的处理步骤。传输只是开始，接下来就是存储。没有存储，就谈不上计算、分析。

那么，为了让5G发展更顺利，目前的存储技术要做出哪些改变？为此，科技日报记者采访了业内相关专家。

存储系统难以匹配传输速度

目前，大型的网络存储系统主要有两类：集中式存储系统和分布式存储系统。其中，集中式存储系统，通常会采用1台存储服务器集群来存放所有数据，这种存储方式的安全性和可靠性较高，但对机房环境的要求高，难以满足大规模并发存储应用的需要。

多合独立的设备上，其采用可扩展的系统结构，利用多台存储服务器分担存储负荷。这种方式不但提高了存储系统的扩展性和存取效率，而且大幅降低了存储系统对硬件设备的需求量。

此外，目前还有一种应用较广的存储模式——“集中式存储+云存储”模式。“这种存储方式综合运用两种存储模式，部分存储任务采用集中式存储模式，工作人员需要建立1个大数据库，同时要配备专属服务器；部分数据存储任务采用云存储模式，把数据存放在由第三方托管的多台虚拟服务器，而非专属的服务器上。”北京工业大学信息与通信工程学院教授张延华在接受科技日报记者采访时说。

面对超高速5G，如今的存储系统准备好了吗？“目前的网络存储设备与5G超快的传输速度，显然是不匹配的。”张延华说，这种不匹配主要体现在两点：在速率上，现有网络存储速度跟不上5G的传输速度；在时延上，5G的传输时延由4G的10毫秒级升至1毫秒级，但部分存储系统的传输时延依旧停留在4G水平。

张延华介绍道，这些不匹配，可能会给用户使用5G带来不便。

首先是容量。张延华解释称，比如，用户在使用5G时，通常要存储下行数据，这些数据会存储在运营商数据中心或用户终端中，传输而来的海量数据，会使数据中心或用户终端承受

巨大的容量压力。

其次是性能。“大型游戏是5G的重要应用之一，数据量剧增，就需要用户的游戏机（电脑端）内存必须随之增加，这也会加大用户在设备上的投入。”张延华说。

需新技术及时“补位登场”

虚拟现实、无人驾驶、高清直播、远程医疗等5G应用，都要求存储系统具有较高的性能和较低的时延。网络存储系统必须跟上业务发展的脚步，才能为用户提供更好的体验。

为迎接5G时代的到来，如今的网络存储系统，需要做出哪些改变？

在张延华看来，新技术“补位登场”，是一条可行的存储升级路径。“除应用相对较多的集中式、分布式存储技术，最近几年，还出现了边缘存储、底层存储等新兴技术，它们都能为数据中心扩充存储容量。”他说。

以边缘存储为例，张延华介绍道，该技术是将数据从远端距离的云服务器端，迁移到离用户数据更近的边缘存储设备端。这种方式可缩短数据传输的时间，降低交互时延，带宽成本以及设备端电能消耗。

“如今，用户数据量正迅速从PB级向EB级攀升，巨大的增量不仅需要扩充容量，还需要能为存储性能带来线性增长的存储架构。”张延华表示，如今，来自视频、网盘、数据库等不同业务

的多种数据，通常会被“一锅烩”式地运行在不同架构的存储系统上。如果架构未能及时升级，就可能在这段时间内“拖累”5G应用的普及。

张延华以5G应用之一自动驾驶为例解释称，比如，现有车联网数据的存储架构，就可能和导航等其他设备数据的存储架构不相容，这或会造成应用普及的障碍。

在张延华看来，未来，随着5G应用的增多，数据类型可能会更多，这就需要一种新的架构，去打破不同类型应用数据间的屏障，让各类应用能在同一架构上进行存储管理。

“5G应用以及物联网技术的普及，可能会使终端设备激增，甚至有望达万亿级。如何才能满足万亿级设备的存储需求、实现海量设备的数据迁移，这些也将是我们后续会面临的问题。”张延华坦言，不过上述问题会随着技术的进步，以及5G应用及设备的成熟，逐步得到有效的解决，“存储系统一定会跟上5G的脚步的”。



热点追踪

“微盘”诈骗卷土重来 从源头防治才能“断根”

兰天鸣 胡洁菲

宣称“10元投资炒白银，5秒开户做沥青”“买涨买跌都能赚”“一部手机创业，一根手指赚钱”……此前，因涉嫌赌博、诈骗等违法犯罪行为，号称“微盘”的虚假投资交易平台遭到证监会等监管部门重点清理打击。记者近日调查发现，当前又有不少虚假投资交易平台利用二维码、APP等悄然“死灰复燃”。受骗人可能遭受何种损失？此类骗局何以屡禁不绝？监管治理如何“断根”？

“成功人士”教投资引用户进圈套

今年5月，上海刘女士通过微信添加了一名自称知道“稳赚不赔的期货投资平台”的“成功人士”为好友。在该“好友”“指导”下，她下载了一个“期货交易APP”，并分三次共向APP账户里充值6666元。照指示，刘女士简单操作了4笔交易，APP内账面立即显示她赚了4000元。在“成功人士”鼓动下，她很快又向APP内账户充值了13万元，进行操作后发现自己APP内账户只剩下13186元。而此时“成功人士”仍怂恿她继续加码充值，投资翻本。

近日，上海市普陀区人民检察院以涉嫌诈骗罪对这起“虚假投资APP诈骗案”犯罪嫌疑人郑某某等20人批准逮捕。案中受害人被诱导使用的虚假期货交易APP，正是典型的“微盘”诈骗手段之一。

据了解，该团伙四名组长手下共有业务员15人。由“经理”提供APP下载二维码并传授员工“话术”。组长负责分配诈骗指标，帮助客户操作虚假期货交易。业务员则负责聊天沟通，吸引客户不断“投资”。

“我们有任务每天发5条朋友圈，加30至50个好友，特别是要在朋友圈把自己包装为成功人士，引人注目，并进行适当的感情联络，引人上钩。”团伙中一名业务员告诉记者。

记者还发现，虽然“微盘”已被监管机构定性为涉嫌违法犯罪行为，并明令禁止，但百度贴吧中仍存在所谓“微盘招商吧”，且关注人数与帖子量均巨大。其中存在大量与“微盘”和建立“微盘”技术公司相关的推介信息，甚至还有传授做“微盘”如何盈利的“忽悠帖”。其中不少帖子将“微盘”诈骗称为“一条轻松赚票子的路”，宣称“做了两年，日入2000”“直推客户可以享受60%的手续返还”。

“病灶”难除致“微盘”诈骗死灰复燃

多名专家告诉记者，“微盘”诈骗死灰复燃的原因在于其“病灶”难除。“以小博大”诱惑大。艾利艾智库高级分析师高婷婷表示，非法期货交易平台等“微盘”诈骗工具开发运营成本低，却利润巨大，刺激“病灶”反复发作。如湖州德清警方“6.23”特大炒原油期货网络诈骗案，涉案金额超1亿元；绍兴越城警方端掉一个期货交易诈骗平台，涉案金额已达1.7亿元。但这些犯罪的原始成本却均在几千元、万余元左右，与之相比，违法收益均在万倍左右。

一些互联网平台管理粗放，缺乏自制，导致相关违法犯罪所需软硬件、模式方法等在互联网上轻易可得。记者通过百度贴吧找到一名为“专业搭建备用”的不法建盘者。他告诉记者，自己“博彩、彩票、期货盘都能做”，建一个期货盘只要6000元，涨幅多少可后台控制，投资者汇入的钱会直接打到后台设置的银行卡里。

“几千块买个系统和服务器，拉一帮业务员就能开个‘微盘’行骗。”高婷婷说，相关违法犯罪活动的组织运作模式均可在互联网上找到，违法犯罪活动被批量复制。这也导致被打击后，不法分子也能轻易将骗局改头换面、“借尸还魂”。

隐蔽性强，查处难度较大。百度贴吧上的不法建盘者还告诉记者，制作虚假期货APP“不存在被追查的风险”，因为“服务器都设在海外”。高婷婷称，虚假期货APP等“微盘”诈骗行为，传播相对隐蔽，“若无人举报，监管部门较难发现。”



图片来源于网络

加强多方联动探索有效防控对策

针对当前“微盘”诈骗APP只要“一串网址”“一个二维码”就能安装、“换个平台就能复活”的“野草式”生存现状，专家建议，工信、网信、金融等部门应加强联动，持续进行源头打击，坚持源头预防，探索有效防控对策。

联动各平台打击预防，让其无处可藏。多名专家表示，搜索引擎、社交平台等仍是“微盘”诈骗相关信息的主要传播渠道。各类网络平台需完善内容审核机制，主动排查举报和处置违法信息；金融监管部门要探索研究虚假期货交易APP绑定银行账户的流水进出规律，对相关支付平台交易额度、频次、跨国资金流动情况加强预警和排查。

手机系统提供方应加强自查。中国人民大学法学院教授刘俊海表示，手机应用商店要加强审查力度，对非法交易为目的的APP要禁止上架；针对由手机系统提供方授权、无需进入应用商店即可用二维码或链接下载的测试版APP，手机系统提供方应切实承担起有关责任，不能让违法犯罪的APP不断“换马甲”“转世投胎”，侵害投资者。

加强投资者教育，让其无利可图。上海金融与法律研究院研究员、执行院长傅蔚冈认为，金融监管部门应加强有效引导，定期公布典型案例，向投资人揭示“微盘”诈骗的特征与危害，及时预警；还应引导投资者理性投资，对自身投资风险承担能力合理评估。专家提醒，一旦误入非法“微盘”交易，应尽快进行举报或报警，尽量将损失降到最低。

(据新华社)