

浅谈信息化战争对大数据存储与分析的要求及对策

孙峥皓, 汪宏昇, 阎岩, 岑小锋, 邓志均

(中国运载火箭技术研究院研发中心, 北京 100076)

摘要: 信息化战争中, 信息将成为新的战略制高点, 随着战场环境的复杂化、作战元素的多元化、战场态势的瞬变化, 海量战场数据存储与分析问题已经成为决定战争胜负的关键问题。本文探讨了信息化条件下对大数据存储与分析的要求, 并提出了几项提高大数据存储与分析能力的对策。

关键词: 信息化战争, 大数据存储与分析, 要求, 对策

0 引言

信息化战争是以大量运用信息技术和信息化武器装备为基础, 以网络化、数字化战场为依托, 以有效地利用信息和争夺信息优势为主要特征的战争^[1]。战场环境的复杂性, 体系对抗的严峻性, 海量数据的实时性, 对体系指挥控制系统的大数据分析与存储能力提出了诸多新的挑战。因此, 如何对海量战场数据进行存储与深度分析, 如何判别数据的真实性, 数据来源的可靠性, 确保数据传输的安全性, 是信息化战争交战双方必须要解决的核心问题。

随着云时代的来临, 大数据正吸引着人们越来越多的关注, 有人预测, “大数据”将是继云计算、物联网之后 IT 产业的又一次颠覆性的技术变革。当前伴随着计算机网络技术的迅速发展和互联网的高速普及, 信息数据量已由 TB (1TB=1024GB) 级升至 PB (1PB=1024TB)、EB (1EB=1024PB)、ZB (1ZB=1024EB) 级, 并仍在呈爆炸式地增长。全球在 2010 年正式进入 ZB 时代, 2012 年全球数据量达到 5.2ZB, 预计到 2020 年, 全球将总共拥有 35ZB 的数据量^[2]。如果把 35ZB 的数据全部刻录到容量为 9GB 的光盘上, 其叠加的高度将达到 233 万公里, 相当于在地球与月球之间往返三次。日益庞大的数据量使得如何有效利用数据, 如何从数据中发掘相关关系等问题成了产业发展的关键。

1. 信息化战争面临的挑战

信息化战争的主要特点就是作战手段信息化、作战空间多维化、作战形式非接触式非线性化^[3]。以计算机为核心的通信技术和网络技术的飞速发展及其在军事领域的广泛应用, 是信息化战争产生的主要推动力。由于信息流在战争中起的作用越来越大, 信息化战争中决定胜负的关键因素已不是消灭敌人有生力量的多少, 而是能否首先抢占制信息权; 战争的对抗形式已不再是武器对武器、平台对平台的简单对抗, 而是由侦察监视系统、指挥控制系统、火力打击系统和综合保障系统等各种作战系统在陆、海、空、天、电(磁)等多维空间上进行的军事作战体系之间的对抗^[4]。

在未来信息化战争中, 体系对抗将在更大程度上依靠各种军事信息系统、软件和数据, 以及在此基础上将多种系统进行综合设计、综合集成、诸军兵种综合运用, 形成以网络为中心的整体信息作战优势, 从而实现各作战要素与武器平台之间的互联互通和互操作, 以确保在正确的时间、正确的地点为正确的决策、指挥和控制提供正确的信息。使信息实现快速合理分发的前提, 是要有对数据的统一调度和管理, 让最即时的战场信息传递到最迫切需要的部门。伴随着战场数据量的增大, 如何高效存储与分析海量数据, 如何从数据中发掘敌我态势的变化, 预测出最合理的作战方案, 使海量数据更好的为信息化战争服务是我们要解决的关键问题。

2 大数据带来的思维变革

当前我们已经进入大数据时代, 如何采用先进的计算机技术来管理数据增长, 将数据转化为可操作的

信息，并驾驭这些信息，使得海量信息能够更好的为信息化战争服务是我们需要思考的核心问题。“大数据”其本质是信息爆炸时代对数据核心价值的再挖掘，实质上是计算机服务时代的来临，是对数据的抽丝剥茧，是从数据中挖掘其内在的灵魂。

这里大数据的“大”不仅仅是数据的大量化，数据类型的多样化，数据流转的动态化，同时也是一种使用数据方式的思维变革。自从18世纪统计学诞生以来，人们已经对数据的分析形成了一种思维惯式，用随机的样本来反映大量数据的特征，这样造成的局面是海量数据的细节信息被放弃了，而抽样统计的结果也不具备拓展性。伴随着云存储、高性能计算等新型计算机技术的高速发展，使人们处理能够收集的全部数据信息成为可能，也使人们有了更多的方法和手段去从海量的数据中发掘其价值，探索其奥秘。大数据的思路即是用能够收集的所有数据来反映事物的细节特征信息，从全部数据中发掘其内涵。

大数据的思维变革，是指人们不再认为数据是静止和陈旧的，大数据开启了一次重大的时代转型，就像望远镜让我们能够感受宇宙，显微镜让我们能够观测微生物一样，大数据正在改变我们的生活以及理解时间的方式，成为新发明和新服务的源泉，而更多的改变正蓄势待发。大数据的核心是预测，它通常被视为人工智能的一部分，或者更确切的说，被视为一种机器学习，但是大数据不是要教机器像人一样思考，相反，它是把数据算法运用到海量的数据上来预测事情发生的可能性。例如朝鲜每次试射导弹，美军都能对朝鲜要发射的时间窗口进行精确估计，根源就在于美军的上百颗侦察卫星，实时监视着地球上的每一个角落，每天存储的信息足以让美军预测出朝鲜军方的可能动向。

大数据的精髓在于我们分析信息时的三个转变，这些转变将改变我们理解和处理问题的方法。第一个转变就是，在大数据时代，我们可以分析更多的数据，有时候甚至可以处理和某个特别现象相关的所有数据，而不再依赖于随机采样。第二个转变就是，研究数据如此之多，以至于我们不再热衷于追求精确度，当我们拥有海量即时数据时，绝对的精准不再是我们追求的目标。第三个转变因前两个转变而促成，即我们不再热衷于寻找因果关系。当处理的数据足够多时，我们更关注于通过如此多的数据可以做什么，而不会关注得出分析结果的原因。例如我们对一种元器件生产周期的全部数据进行记录，通过分析这个生产批次的所有元器件的数据，我们就能发现这个元器件有哪些薄弱环节，会分析出可能的安全隐患，但是我们不会去关注是什么原因造成这个元器件的可能缺陷，只会去关注如何避免其出现故障。

大数据的科学价值和社会价值正是体现在这里，一方面，对大数据的掌握程度可以转化为经济价值的来源。另一方面，大数据已经撼动了世界的方方面面，从商业科技到军事、医疗、政府、教育、经济、人文以及社会的其他各个领域。

3 信息化战争对大数据处理的要求

信息化战争中，体系对抗特征明显，多维一体协同作战成为作战的基本样式。在每一次战争中情报、监视和侦察系统产生了海量的图像数据、电子侦察数据、可见光谱数据、情报数据、战场实时态势数据以及后勤保障、军队调动等指令数据。在这样的背景下，信息化战争对数据分析与存储产生了特殊的要求。

(1) 处理实时性要求。信息化战争的作战理念是“网络中心战”，强调的是以网络为中心，将分散在各地的部队、探测器、武器系统和作战平台融为一体，实现信息共享、战场透明，兵力分散、火力集中，快速反应、整体联动，以便对敌实施快速、精确和连续的打击。信息化战争中产生的海量数据最终要汇集到各级指挥中心和处理中心，这些数据源源不断的从各种传感器、情报机构以及信息中心传输汇聚到一起，要对这些数据进行实时的处理，对数据处理系统的处理能力和处理的实时性要求很高。

(2) 统一调度和高效管理要求。海量数据的来源不同，对数据存储的要求也不相同。一方面要求数据存储系统提供足够的存储空间，提供足够可靠的数据存储，确保任何情况下数据可以按需存入和取出；另一方面，要求在存储的同时，能够对数据进行分类和分析，结构化的数据、非结构化的数据，图像数据、图形数据、雷达数据、导航数据、指控数据等等，能够按照要求对数据进行分类存储，同时对数据进行分析与挖掘，使得宝贵的数据资源中的信息能够充分利用。

(3) 数据安全的要求。信息化战争对数据源的信息安全要求体现在确保数据源是真实的、己方的，这

里主要解决身份识别的问题,即信息源的身份认证问题,防止敌方信息对抗系统对己方信息系统注入虚假的欺骗信息^[5]。信息化战争对数据传输的信息安全要求体现在解决防窃听、抗干扰和防止虚假信息欺骗的问题,即在复杂的战场对抗环境下,防止敌方信息对抗系统窃听己方的信息传输内容、对己方系统信号的干扰和信息系统注入虚假的欺骗信息^[6]。信息化战争对数据应用的信息安全要求体现在对信息使用者访问权限的控制,对信息使用者身份的识别。

(4) 弹性扩容的要求。大数据时代的数据量是呈几何指数关系递增的,如何构建未来的指挥控制系统,使其具备可伸缩的弹性服务是需要解决的核心难题。传统的采用物理服务器的方式扩展性差,无法实现对数据的统一管理,资源利用率低,部分服务器处于闲置状态;数据的扩展性依靠的是存储设备的垂直拓展,造价昂贵、可维护性差,缺乏弹性拓展,灵活性差。建立基于云计算技术的指挥控制系统可以很好的解决上述这些问题。弹性的指挥控制系统可以实现资源的按需供给和动态管理,弹性的主要特征是使系统能够可大可小、可增可减的利用计算资源。弹性的主要目的是不必担心资源的过渡供给导致额外的使用开销,亦不必担心资源的供给不足导致应用程序不能很好的运行,所有资源以自适应伸缩的方式供应。

4 提高大数据处理能力的几点对策

随着信息化战争的发展,对海量战场数据的存储与分析提出了诸多要求,针对上述问题,国内外已经开展了很多研究,并取得了一定的成果,下面主要讨论一些目前比较前沿,并且有潜力解决上述问题的相关技术。

(1) 大数据并行处理

大数据可以通过 MapReduce 这一并行处理技术来提高数据的处理速度。采用 MapReduce 模型,将要执行的问题分解成 Map (映射) 和 Reduce (化简) 的方式,先通过 Map 程序将数据切割成不相关的区块,分配(调度)给大量计算机处理,达到分布式运算的效果,再通过 Reduce 程序将结果汇整输出^[7]。MapReduce 的设计初衷是通过大量廉价服务器实现大数据并行处理,对数据一致性要求不高,其突出优势是具有拓展性和可用性,特别适用于海量的结构化、半结构化及非结构化数据的混合处理。MapReduce 将传统的查询、分解及数据分析进行分布式处理,将处理任务分配到不同的处理节点,因此具有更强的并行处理能力。

(2) 内存计算

随着大数据时代的来临,内存计算的时代也即将到来。很多人都在说:“不管磁盘是否消融,闪存都是将来的一个趋势。”内存计算,实质上就是 CPU 直接从内存而非硬盘上读取数据,并对数据进行计算、分析。此项技术是对传统数据处理方式的一种加速,是实施海量数据分析的关键技术。由于服务器在处理数据时,CPU 首先会从缓存中找数据,缓存中找不到,再从内存中找,内存里没有,再从硬盘上读取。在传统数据库查询中,磁盘访问时间是一个主要的瓶颈。研究发现,如果让查询在读写速度快很多倍的内存中进行,而不用访问物理磁盘,将会大大提升处理性能。通过举一个例子来说明内存计算可以快到什么样的程度,它可以做到 30 亿次的扫描/秒/核,1250 万次的聚合/秒/核,150 万次的插入/秒,250TB/小时的数据处理,1 亿表单/小时。

(3) NoSQL 数据库

随着大数据时代的来临,海量数据的存储给数据库提出了很高的并发负载要求,当前的数据存储往往要面对每秒上万次的读写速度,传统的关系型数据库面对上万次查询还勉强顶得住,但是应付上万次写数据请求,硬盘 IO 就已经无法承受。NoSQL 数据库是一种建立在云平台基础上的新型数据处理模式。NoSQL 在很多情况下又叫做云数据库,由于其处理数据的模式完全是分布于各种低成本服务器和存储磁盘,因此它可以帮助各种交互性应用快速处理过程中的海量数据。正常的数据库需要将数据进行归类组织,类似于姓名和账号这些数据需要进行结构化和标签化,但是 NoSQL 数据库完全不关心这些,它能处理各种类型的文档,支持分布式存储,能透明地扩展节点。

5. 结束语

伴随着信息化战争的高速发展, 指挥控制系统, 天基支持系统, 信息处理系统, 各种侦察、监视、探测系统的信息量越来越大, 数据在指挥机构中的管理也变得空前重要和复杂, 对资源和安全管理工作也提出了全新的挑战。本文提出了几点信息化战争对大数据存储与分析的要求, 也提出了几点有潜力解决上述问题的相关技术。这些技术的突破有望使指挥控制系统、信息传输系统以及军队作战方式、战场形态等发生新的重大变化, 在军队信息化建设和信息作战中具有广阔的应用前景。

参考文献:

- [1] 罗雪山等. 军事信息系统体系结构技术[M]. 北京: 国防工业出版社, 2010.
- [2] 孟薇薇等. 信息爆炸时代的新概念-大数据[J]. 2012.
- [3] 李春祥, 赵琦. 浅谈信息化战争条件下高素质指挥人才的培养[J], 中国系统建模与仿真技术高层论坛论文集, 2008.
- [4] 徐小岩. 军队信息化建设概率[M], 北京: 解放军出版社, 2005.
- [5] 杨健, 汪海航, 王剑等. 云计算安全问题研究综述[J]. 小型微型计算机系统, 2012, 33(3): 472-479.
- [6] D.Stucki, N.Gisin, O.Guinnard, et al. Quantum key distribution over 67km with a plug&play system [J]. New Journal of Physics, 2002, 4(41): 1-8.
- [7] Mell P, Grance T. The NIST definition of cloud computing [R]. National Institute of Standards and Technology, 2011.