

# 大数据在部队信息资源管理中的应用

尹承督<sup>1,2</sup>, 贾红丽<sup>1</sup>, 李卫东<sup>2</sup>

(1. 军械工程学院 装备指挥与管理系, 石家庄 050003; 2. 77606 部队, 拉萨 850000)

**摘要:**针对部队管理数据日益出现的多源异构、数据量剧增的现象,基于大数据处理工具和技术,分析了大数据系统在部队信息资源管理中的运用与数据交汇中心的构建以及关键技术的处理。

**关键词:**大数据;大数据分析技术;信息资源;大数据分析系统

**本文引用格式:**尹承督,贾红丽,李卫东.大数据在部队信息资源管理中的应用[J].兵器装备工程学报,2016(8):118-121.

**Citation format:**YIN Cheng-du, JIA Hong-li, LI Wei-dong. Application of Big Data in Military Information Resources Management[J]. Journal of Ordnance Equipment Engineering, 2016(8):118-121.

中图分类号:E737

文献标识码:A

文章编号:2096-2304(2016)08-0118-04

## Application of Big Data in Military Information Resources Management

YIN Cheng-du<sup>1,2</sup>, JIA Hong-li<sup>1</sup>, LI Wei-dong<sup>2</sup>

(1. Department of Equipment Command and Administration, Ordnance Engineering College, Shijiazhuang 050003, China; 2. The No. 77606<sup>th</sup> Troop of PLA, Lhasa 850000, China)

**Abstract:** Aiming at the phenomenon that the multi-source heterogeneous data appearing increasingly and data is increasing quickly about military management data, this paper mainly analyzed the application of big data system in army information resources management and the construction of data interchange center as well as the disposing of key technology based on the processing tools and techniques of big data.

**Key words:** big data; big data analysis technology; information resources; big data analysis system

当前部队信息化建设不断发展,信息化管理平台广泛应用,装备信息化程度逐步提高,数据常态化机制建立,军事数据呈现出多源异构、数量庞大等“大数据”现象,部队信息资源逐渐成为现代作战的重要因素,如何有效管理、分析部队数据资源,从日益复杂庞大的数据信息中得到支持战争的优化决策,对数据进一步“提纯”成为部队发展的重要研究课题。

### 1 大数据概念及大数据系统

“大数据”是从英语“Big Date”一词翻译过来的,大数据的概念并非刚刚出现,在20世纪90年代初,数据库之父比

尔·恩门就多次提及“Big Date”,大数据的定义至今没有得到统一。

#### 1.1 大数据定义与特点

大数据有多种定义,维基百科、麦肯锡、亚马逊约翰·罗瑟、Infomatica中国首席顾问但彬等都都对大数据进行过定义<sup>[1]</sup>。综合当今研究状况,大数据是指其大小超过了典型数据库软件的采集、存储、管理、分析等能力的数据集。

信息产业界通常用4个V概括大数据的特征,即体量大(Volume),一般指数据量在10TB以上的数据。类型多(Variety),数据来源种类丰富,包括结构化数据、半结构化数据、准结构化数据、非结构化数据。速度快(Velocity),一是数据增长量快;二是数据处理速度快,数据处理遵循“1秒”定律。

收稿日期:2-16-02-18;修回日期:2016-03-16

基金项目:全军军事类研究生资助课题

作者简介:尹承督(1984—),男,硕士研究生,助理工程师,主要从事装备建设与发展研究;贾红丽(1972—),女,副教授,博士,硕士生导师,主要从事装备建设与发展研究。

价值性(Value),一是价值密度低;二是潜在价值高,大数据处理技术的迅速发展,促使在海量繁杂的信息中进行“沙中捞金”,获得高价值的知识与决策<sup>[2]</sup>。

## 1.2 大数据处理工具 Hadoop 系统

由于 Hadoop 处理平台是开源性操作平台,且出现的比较早,使用的比较成熟,而且 Hadoop 处理平台适用于廉价的服务器集群,数据存储与服务仅仅分布在 HDFS 与 HBase 两个层次,从而大大减少资源的浪费<sup>[3]</sup>。且由于中国电信、移动、阿里、百度等知名企业所应用的都是 Hadoop 系统,对于我军使用该系统有很好的军民融合的道路途径,便于部队与地方对非保密性关键部件进行合作开发。

### 1) Hadoop 系统介绍

2003 年 Google 连续发表了 GFS、MapReduce、BigTable 的论文,此时开源搜索引擎 Nutch 和开源全文检索 Lucene 的作者 Doug Cutting 从 Google 发表的文章中找到提高平台性能和可靠性的方法。从 2004 年,他用两年的时间开发了深受广大欢迎的 Hadoop 系统。Hadoop 目前是 Apache 组织所推进的项目,该项目由两部分组成,一个是基础部分,另一个是配套部分,如图 1 所示。

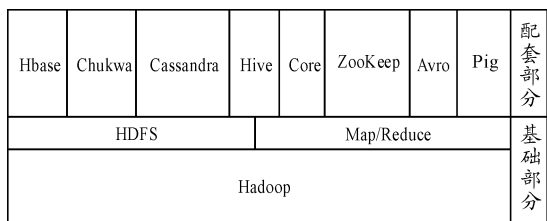


图 1 Hadoop 基础构架

### 2) Hadoop 系统数据处理技术介绍

Hadoop 系统的低成本、高容错性、高扩展性且无需预先定义模式,被广泛用于搭建大数据系统,特别适于精处理与分析数据,有效保存原始数据,在模式识别、索引编制、推荐引擎建立和分析方面都广泛使用<sup>[4]</sup>。下面简单介绍 Hadoop 系统的 3 个关键技术。

分布式文件系统 HDFS,主要运行在廉价的硬件集群上,可存储大规模数据集,进行流式数据读写(一次写入,多次读取),便于集群规模动态扩展(错误检测与快速自动恢复),数据处理时采用的是移动计算不是移动数据。用块可简化存储子系统,适合于提供容错与复制操作。HDFS 有名称节点 NameNode 与数据节点 DataNode 两类节点,以主从模式运行,主控服务器一般只有一个,从控服务器主要是防止名称节点损坏,复制元数据,因此设置二级名称节点。

分布式编程模式 Map/Reduce,可以将 Map 函数看成“映射-分析数据”的过程,将 Reduce 函数看成“化简-归纳数据”的过程。对于大数据(运算对象分布)拆分成多个 Map(运算能力分布)可处理的分任务,并传递到远端数据节点进行运行,最后由 Reduce 进行合并得到综合结果,其核心思想

“分而治之、移动逻辑、屏蔽底层、处理定制”,这就是 Hadoop 系统对大数据分析处理的一般过程。

面向列存储的分布式数据库系统 HBase, HBase 不是关系型数据库,而是常说的 NOSQL 数据库,它可以以线性方式扩展存储规模,而且是动态扩展,无需停机扩展,避免了关系型数据库扩展的安装、维护复杂,受制于节点,原代码需要更换等缺点,实现大规模数据的存储<sup>[5]</sup>。

## 2 部队信息资源管理现状

随着部队信息化建设不断深入发展,各种信息化管理平台建立与广泛使用,部队时刻都在产生大量的数据信息,这为部队建立了一个规模巨大,具有大数据特征的数据信息资源库。面对部队大量信息资源,部队在如何分析、整合、梳理、提炼这些信息资源的过程中,还存在很多问题。总体表现缺乏顶层设计、信息化基础设施建设不够完善、软件系统兼容性不好、运行机制不健全、信息化人才匮乏<sup>[6]</sup>。

1) 顶层设计缺乏指导性,信息管理涉及部队不同层面,是一项复杂的系统工程,顶层设计要充分考虑我军信息化水平与发展条件,避免造成盲目开发,自行其是,重复建设等现象。

2) 思维模式落后,传统的被动反应式管理很难适应部队精细化管理。随着部队信息化建设深入发展,必须加强对信息资源的采集、传输、处理,重视对部队信息资源的快速、深度利用的程度。

3) 信息管理系统类型多,相互之间信息不流通,数据冗余问题突出,并且对部队业务工作指导性欠缺,缺乏信息管理人才,部队应用这些信息系统的积极性不高,甚至有些系统基本处于闲置状态,处于自生自灭的尴尬境地,难以发挥应有的效能。

4) 由总部及军区开发的系统多,虽说这些信息系统给工作带来一定的便利,使很多部门管理工作有了信息化辅助工具,数据查询方便高效,但是这些信息系统的开发相对独立,没有充分考虑各系统之间的功能衔接和数据流通,经常出现“信息孤岛”,需要反复输入相关信息,容易造成数据的冗余与不一致。加大了数据输入人员的工作量,并且需要的数据信息统计不上来,而不需要的信息泛滥成灾。

如今部队面临着大范围的体制编制的变革,正值部队数据重建的大好时机,采用适合于部队信息化发展的数据管理平台,构建未来数据多种类、数据量大的大数据处理平台,在最大程度上发挥数据决策分析工作的支持作用,是当今部队新型军事变革必须重视的一项重大工程。

## 3 部队信息资源管理中的大数据分析系统构建

随着信息技术的发展,人类社会已经进入一个数据爆炸

的大数据时代,大数据对人们生活产生巨大影响同时,也对部队信息化发展带来巨大影响,尤其是部队信息资源出现多源异构,数据量剧增的现象,建立适应部队信息资源管理的大数据分析系统,实现由数据向战斗力的转化,对提升部队信息化作战能力有巨大推进作用。

### 3.1 大数据分析系统示意图

由于部队信息化管理平台的广泛使用,装备信息化程度提高,各种智能化、自动化设备的使用,使得部队各种信息量成倍增长。相关数据呈现多样化,视频、音频、图像、表格等多种形式的信息不断大量产生。虽然从单一的部门来看数据量还达不到“大数据”的规模,但是信息化条件下要求联合作战能力的提升,所涉及的装备复杂多样,单位数量多,而且战场环境、战场态势、作战工程等的多样性,要做到“能打仗,打胜仗”,各单位的数据信息必然要求上下互联,纵横互通,这样综合看待数据量,尤其是多兵种、大数据量、样式多样化的数据,要从中找到支持决策的信息犹如大海捞针,传统的分析已严重不能应对部队海量多源异构的信息资源处理。而且大数据在部队的个别领域已经呈现,例如:航空、航天、情报侦察、军事气象等领域,数据规模达到PB级别<sup>[7]</sup>,因此部队信息资源数据需要从大局上以大数据来对待,在部队数据管理、分析、处理等方面对大数据技术提出具体要求。

以PC1—PC9代表不同单位、不同装备、不同作战指挥部、不同的作战任务单元,以Hadoop系统建立信息交汇中心,将整个部队信息资源体系构成以信息交汇中心为基础的数据处理网络,如图2所示。海量多源异构的信息资源的汇聚,从作战全局、管理部署、任务决策出发,经过数据分析产生相关的决策信息迅速反馈给各个作战单元,支持各单位做出相应的决策部署。

部队在联合作战中更要重视从大量的数据中挖掘有用的价值,尤其是对来自不同单位,同时发生的数据进行对比关联,得出及时的知识与决策,从而为指挥官准确掌握战场态势、情报信息,及时制定作战决策。

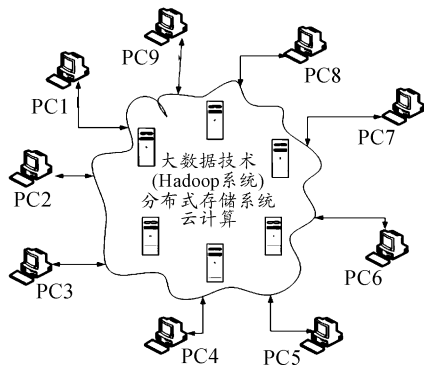


图2 部队基于大数据分析的数据中心网络建立示意图

### 3.2 数据交汇中心的建立及关键技术

针对部队结构化数据与非结构化数据并存的现象,研究

开发支持结构化与非结构化数据统一管理的大数据管理平台,成为一项非常迫切的任务,因此对部队信息资源的管理,采用大数据系统,其数据交汇中心的建立至关重要,如图3所示。

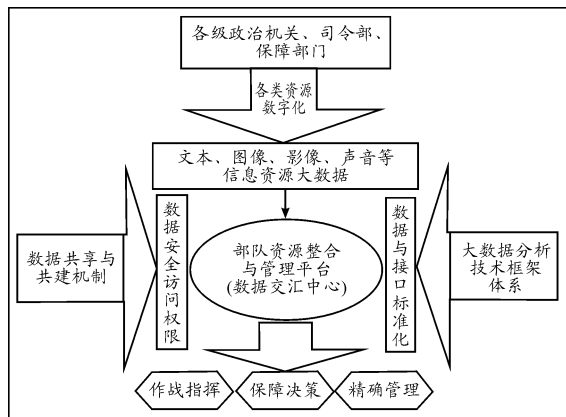


图3 部队信息资源交汇中心建立示意图

需要解决的主要技术与问题:

1) 标准化,这里包括信息资源数字化标准和大数据存储与分析技术应用接口标准化。统一部队信息资源的数字化标准,尤其是多源异构数据格式的统一性,并且在面临部队大面积改革,隶属关系重大变动时机,大量数据进行数据格式统一更有利于部队信息资源的共享,并且统一部队大数据分析系统,统一采用Hadoop系统,对部队联合作战的数据上下顺畅,同级之间资源信息共享具有重大意义。大数据存储分析技术应用接口的标准化,对于提高数据共享同样具有重要意义。

2) 加强对部队信息资源大数据分析的基础技术利用。Hadoop大数据系统的分布式编程模式Map/Reduce的使用,采用任务分解、归纳提高数据分析的速率,而且大大降低对硬件的要求,利于节约部队硬件使用成本。Hadoop大数据系统的HBase动态数据库的使用,有利于随着数据的增加在不停机的同时随机扩充部队数据存储量,提高数据分析的不间断,持续性,有利于部队对信息掌握及时性,提高决策的时效性。

3) 注重数据安全访问机制,数据的安全与防护不容忽视,确保数据的详实、安全、可靠、可信,防止丢失、损坏、篡改、删除等恶性事件发生。数据安全与防护设置不当将会对我军发展造成不可估量的损失。设置相应的数据使用访问权限,也是部队大数据分析系统确保数据安全可靠必须高度重视的技术。

4) 所有技术的发展与应用都必须重视人的作用,相应的大数据分析系统的使用在技术发展的同时,必须加强信息化人才的培养,确保技术与人才同时发挥合力,使大数据系统在部队发挥最大效益,促进部队信息资源真正实现由数据到知识的转变。

## 4 结束语

大数据时代的到来,预示着一场新的军事革命的即将来临,伴随部队信息化建设的发展,信息化管理平台在各部门的广泛使用,使得数据信息成为部队战斗力生成模式转变的新型催化剂、融合剂。由于数据的多结构,数据量的剧增,仅靠传统的数据分析,必然造成数据利用的局限性,时效性差,为了更好地利用部队数据,实现数据由辅助决策到支持决策,真正做到“数据到决策”的转变,必须寻求更适于多源异构的大数据分析手段,因此大数据分析技术成为部队未来的必然选择。同时在面临我军体制编制大变革的关键时期,构建适用于我军发展的大数据系统平台,对部队信息化的发展也具有强大的推动作用。

## 参考文献:

- [1] 杨巨龙. 大数据技术全解基础、设计、开发与实践[M]. 北京:电子工业出版社,2014.
- [2] 赵刚. 大数据技术与应用实践指南[M]. 北京:电子工业出版社,2014.
- [3] 李军. 大数据从海量到精准[M]. 北京:清华大学出版社,2015.
- [4] 邢程,李玉梅,胡荣亮.“大数据”下的军队信息资源建设[J]. 教育技术研究,2014(3):65-68.
- [5] 姜念云,张松海,谢夏. 大数据分析技术在文化资源管理中的应用[J]. 中国基础科学,2014(1):17-20,27.
- [6] 王向博. 基于数据挖掘的装备维修信息管理研究[D]. 军械工程学院装备指挥与管理系,2013.
- [7] 涂新莉,刘波,林伟伟. 大数据研究综述[J]. 计算机应用研究,2014,31(6):1612-1616,1623.

(责任编辑 杨继森)

(上接第92页)

- [23] PURANDARE R, DWYER M B, ELBAUM S. Monitor optimization via stutter-equivalent loop transformation [C]//Proceedings of the ACM international conference on Object oriented programming systems languages and applications, ACM; Reno/Tahoe, Nevada, USA. 2010:270-285.
- [24] DIEP M M, DWYER M B, ELBAUM S. Lattice-Based Sampling for Path Property Monitoring [J]. ACM Trans Softw Eng Methodol, 2011, 21(1):1-43.
- [25] BONAKDARPOUR B, NAVABPOUR S, FISCHMEISTER S. Sampling-based runtime verification [C]//Proceedings of the 17th international conference on Formal methods, Springer-Verlag; Limerick, Ireland. 2011:88-102.
- [26] NAVABPOUR S, WU W W C, BONAKDARPOUR B, et al. Efficient techniques for near-optimal instrumentation in time-triggered runtime verification [C]//Runtime Verification, 2011.
- [27] PIKE L. Copilot: a hard real-time runtime monitor [C]//Proceedings of the First international conference on Runtime verification, Springer-Verlag; St. Julians, Malta. 2010:345-359.
- [28] JIN G. Instrumentation and sampling strategies for cooperative concurrency bug isolation [C]//Proceedings of the ACM international conference on Object oriented programming systems languages and applications, Reno/Tahoe, Nevada, USA. 2010:241-255.
- [29] CALLANAN S. Software monitoring with bounded overhead [C]//Parallel and Distributed Processing, IPDPS 2008.
- [30] DWYER M B, PURANDARE R, PERSON S. Runtime verification in context: can optimizing error detection improve fault diagnosis [C]//Proceedings of the First international conference on Runtime verification, Springer-Verlag; St. Julians, Malta. 2010:36-50.
- [31] JIN D. Garbage collection for monitoring parametric properties [C]//Proceedings of the 32nd ACM SIGPLAN conference on Programming language design and implementation. San Jose, California, USA. 2011:415-424.

(责任编辑 杨继森)