# 支撑指控系统智能化的大数据处理技术研究

周凯，黄治华

（中国船舶重工集团公司第七〇九研究所，武汉，430074）

**摘要**：在指控系统的发展过程中，面临数据规模剧增、新业务应用不断出现所带来的新问题，智能化成为指控系统发展的趋势和必然要求。大数据处理作为当前已经在商用领域得到成熟应用的数据处理技术，为指控系统应对复杂海战场环境下的海量战场信息处理问题提供了有益的参考。本文对目前主流的大数据处理技术和系统架构进行了梳理，并结合指控系统智能化的需求，指出了需要考虑的关键问题，并提出可供参考的实现方法。文中最后提出的支撑指控智能化的数据处理平台构想，为建立海战场复杂环境下的大数据处理平台提供了思路。

**关键词**：大数据；指控系统；智能化

## 1 舰载指控系统智能化概述

随着未来一段时期，海军作战需求变化、作战样式多样、作战范围广阔的发展要求，需要指控系统进一步提升信息处理及综合显示、辅助作战筹划与决策、兵力武器协同指挥等方面的能力，并具备更多的智能化特征，主要包括情报保障综合化、决策支持智能化、指挥控制一体化、人机交互快捷化等。

情报保障综合化要求将原有离散的需要指挥员判断综合的各类态势和信息进行多维综合展示，便于指挥员快速掌握战场态势，做出准确判断；决策支持智能化是指将指挥员的经验和知识转化为具有推理能力和人工智能地辅助决策工具，提供定性定量相结合、人机结合的决策手段，提高指挥决策的效能；指挥控制一体化要求充分整合目前分散孤立的各类指挥与控制功能，减少指挥控制的层次和环节，提高指挥控制的效率；人机交互快捷化是指要进一步增强人机交互的友好和快捷程度，形成从指挥员脑中决策到武器随动的快速反应机制，缩短系统反应时间。

当前指控系统的智能化程度不高，在实现指控系统智能化过程中还面临诸多问题。随着海战场环境日益复杂而导致的作战数据量不断增大，有关数据处理方面的问题也日渐突出。现在大数据技术在商业领域中的应用方兴未艾，其所面向的应用场景与指控系统中所面临的数据处理问题有相通之处，值得我们深入研究、借鉴参考。

## 2 大数据与指控智能化

2008 年 9 月《科学》杂志上发表了一篇文章《Big Data: Science in the Petabyte Era》，“大数据”这个词开始被广泛传播。大数据是数据量增长从量变到质变的里程碑，但也不能简单地以数据规模来界定大数据，而要考虑满足用户需求的数据处理与分析的复杂程度。针对简单的用户需求（如关键字搜索），数据量为 TB 至 PB 级时可称为大数据；而针对复杂的用户需求（如数据挖掘），数据量为 GB 至 TB 级时即可称为大数据[1]。

大数据的特征主要是规模大、类型多、变化快[2]。由于数据的异质异构、无结构及不可信等特征，大数据的管理和分析研究需要解决可表示、可处理和可靠性三个关键问题。[3]在商用领域，其主要技术挑战在于数据的异构性和不完备性、数据处理的时效性、数据的隐私保护和大数据的分析处理等。在“数据科学”领域，大数据管理及处理能力已经成为引领网络时代 IT 产业发展的关键。获取大量真实的运行数据并建立对其进行动态高效处理的能力，将成为产业竞争力的体现。

与商用领域类似，在指控系统的数据处理过程中，其所处理的数据也具有规模大、类型多、变化快等特征；对于数据处理的时效性、数据保护也有较高要求。此外，在数据处理分析的目标上也类似，在商用领域的大数据分析中，其主要目标是为了从大量模糊信息中分析得到精准的用户需求、挖掘利益增长点；而指控系统中的数据分析则是为了提高作战信息的精确度，为作战人员提供快速、准确的辅助决策支撑。

然而，与商用领域相比，指控系统中的数据处理也有自己的特点。如在所处理的数据量上，指控系统中所处理的数据量一般达不到商用领域的数量级（如搜索引擎、大型电子商务应用等），也基本不会面临处理和存储能力跟不上数据量增长的情况；在数据组成等方面也存在差异，指控系统中处理的数据组成通常较为固定，数据中的结构化数据较多、非结构化数据较少。

尽管在具体应用场景中存在上述差异，但在一些基本特征上，指控系统中的数据处理需求与商用领域仍有相通之处，可以借鉴现有商用领域中的大数据处理技术，结合指控系统智能化的需求和自身特点，为解决指控系统数据处理中的问题提供一些技术支撑。

## 3 支撑指控智能化的大数据处理技术

在大数据生态系统中覆盖了大量的技术和实现，其中一些在大数据技术发展道路中起到了巨大的推动作用，下面列举了其中的一些典型代表。

（1）Hadoop。Hadoop 是 Apache 开源组织的一个分布式计算开源框架，凭借其开源和易用的特性，Hadoop 不仅是大数据时代数据处理的首选，也是拥有海量数据处理需求的公司的标准配置。基于 Hadoop 可以轻松地编写可处理海量数据的分布式并行程序，并将其运行于由成百上千个节点组成的大规模计算机集群上，许多商业创新都围绕 Hadoop 展开。

（2）NoSQL。与 Hadoop 密切相关的的 NoSQL 也一直是大数据领域的热点。NoSQL 凭借其高性能和可扩展性等优势，成为关系型数据库的强劲对手，在大数据时代占据一席之地。但 NoSQL 数据库并不是要取代现在广泛应用的传统数据库，而是采用一种非关系型的方式解决数据的存储和计算的问题。

（3）NewSQL。由于 NoSQL 不支持结构化查询语言（SQL），这给开发人员带来诸多不便。为了同时满足高性能和支持 SQL 两个方面，NewSQL 被设计出来。NewSQL 作为全新的关系数据库产品，或将关系模型的优势发挥到分布式体系结构中，或将关系数据库的性能提升到不必进行横向扩展的程度，这使得 NoSQL 面临前所未有的挑战。典型的 NewSQL 有 VoltDB、Marklogic、Xeround、NuoDB 等。

（4）Data Marketplace。除了解决大数据处理、存储问题之外，开放数据资源也在相当大程度上加速了大数据技术的发展。目前大部分的企业所面对的数据都是由内部系统或者交易记录日志之类的东西所产生的，然而如果能够获得企业自己无法获得，或者已经被处理过的外部数据，那么内外数据融合分析后产生的价值将不可估量。

从上面所列的技术或产品中可以看到，它们涉及数据存储、数据处理、数据分析等多个方面，指控系统的数据处理过程也离不开这几方面，以下分别从这三方面讨论可用于支撑指控系统数据处理，提高其智能化程度的相关技术。

### 3.1 数据存储

指控系统中所处理的数据涉及各种报文和接口，数据类型复杂、多变，随着应用规模的增大，随之而来的是数据量的剧增，新数据类型不断涌现，用户需求呈现出多样性，对数据的管理和维护难度大大增加，传统的数据库适用的数据结构、并发控制、故障恢复等技术在新的环境下面临挑战。针对海量异构数据，如何构建一个模型来对其进行规范表达，如何基于该模型来实现对其进行有效存储和高效查询是亟须解决的问题。

以 Oracle、MySQL 等为代表的传统关系型数据库在存储结构化数据上表现出色，通用性较强，但对于半结构化数据、非结构化数据的存储，以及对具有某些特定要求（如海量数据条件下的高效读写、高可用性）的应用则显得并不那么适合。

NoSQL（Not Only SQL）是近年来新兴的非关系型数据库，其特点是自由的 schema、数据多处备份、简单的编程 API、数据的最终一致性等[4]。根据 NoSQL 数据库内部数据组织的不同， NoSQL 大致可分为列存储、文档存储、键值存储、图存储等类型，其中，HBase、MongoDB、Cassandra、Redis 等 NoSQL 数据库已被相当多的企业和开发人员所熟知。目前的 NoSQL 运动正在通过放弃关系型数据库强大的 SQL 查询语言、事务的一致性以及范式的约束，或者采用键值数据格式存储，以获得高效灵活的大数据处理能力。

以 MongDB 为例，它支持的数据结构非常松散，可以存储比较复杂的数据类型，同时支持强大的查询语言，几乎可以实现类似关系数据库单表查询的绝大部分功能，而且还支持对数据库建立索引，适用于海量数据的存储与查询等应用场景；又如 Redis 数据库，它不仅支持简单的键值存储，还支持 list、set、hash 等数据结构的存储，Redis 本质上是一种内存数据库，整个数据库系统加载在内存中操作，定期通过异步操作将数据保存至硬盘，具有很出色的读写性能

表 1 中给出了当前数据库领域中的主要数据库分类及各类数据库的主要实现。

表 1 各种数据库分类

| | | |
|---|---|---|
| 关系型数据库 | | Oracle、MySQL、SQL Server、DB2 等 |
| NoSQL 数据库 | 键值存储（key-value） | Redis、Memcached、Riak、Amazon DynamoDB 等 |
| | 列存储 | Cassandra、HBase、Google BigTable 等 |
| | 文件存储 | MongoDB、CouchDB、RavenDB 等 |
| | 图存储 | Neo4j、InfiniteGraph、DEX 等 |
| | RTF 存储 | Apache Jena、Sesame 等 |
| 其他 | | 本地 XML 数据库、面向对象数据库、内容存储数据库等 |

针对指控系统对复杂数据类型表示、海量数据存储效率和高效读写等方面的需求，可以采用具有针对性的 NoSQL 存储方案，选择适用于特定应用场景的数据库实现。

在商用领域，新型的 NoSQL 数据库已经有了比较成熟的应用，但要应用到指控环境中，还需要结合实际应用、工作量等多方面因素进行综合考虑，形成适合具体应用的方案。因为新的数据存储方案的引入，意味着要对现有数据存储策略、算法等进行修改甚至重新设计，同时还要考虑对已有应用的影响；另一方面，使用新型数据库并不意味着抛弃原有数据库产品和方案，而是为不同的应用场景提供更多的选择，解决问题的重要原则是选择合适的方案处理合适的业务场景，而非一味盲目推动新旧交替。

### 3.2 数据处理平台

在指控系统中，所处理的数据具有量大、分布广等特点，对于数据处理的时效性、准确性等有很高的要求，特别是在敌情侦测分析、火力打击等战时业务场景下，并且随着各种应用的增多，数据增长速度越来越快，这进一步提高了对数据处理效率和性能的要求。提高指挥效率的关键是缩短情报信息处理时间。从情报信息处理速度与情报信息利用率的关系看，处理速度越快，情报信息的利用率就越高，反之则利用率越低。从作战指挥各个环节必用时间看，情报信息处理环节也是提高指挥效率潜力最大的环节[5]。因此，建立应对上述需求的数据处理平台十分必要。

在数据规模极大的情况下，数据具有规模大和分布性等特点，使用传统的数据处理方式处理大数据不论在处理效率还是处理效果上都不能满足应用对数据处理的需求。因此对于大数据的处理，需要有新的处理方式。

硬件已经不再是制约处理能力的第一要素，通过具有一定规模的计算中心，加上设计完整的计算框架，就可以保证不同节点及单个节点不同进程间的协同工作能力，实现可靠、高性能的强大数据处理和分析。使用者只需通过简单而强大的编程框架提交需要完成的计算任务以及相关数据，系统就可以自动安排和处理支撑计算所需的其他复杂工作，如输入数据的分割、中间数据的传输分布、多机环境下的程序执行和调度以及输出数据的聚合等。

当前在商用领域得到广泛应用的数据处理框架 Map/Reduce 是面向海量数据处理应用的有效途径之一，其框架实现可作为指控系统数据处理的有益参考。

Map/Reduce 是 2004 年由谷歌公司提出的一个用来进行并行处理和生成大数据集的框架，它是将并行编程中复杂的业务逻辑进行抽象的结果。它将简单的计算作为接口展现在前端，而对并行化处理、容错、数据分布和负载均衡都进行了隐藏，现在得到广泛应用的 Hadoop 就是其开源实现。一个 MapReduce 作业通常会把输入的数据集切分为若干独立的数据块，由 map 任务以完全并行的方式处理。框架会对 map 的输出进行排序，然后把结果输入给 reduce 任务。整个框架由一个单独的 JobTracker 和每个集群节点中的 TaskTracker 共同组成。JobTracker 负责调度构成一个作业的所有任务，这些任务分布在不同的 TaskTracker

上，JobTracker 监控它们的执行，重新执行已经失败的任务，而 TaskTracker 仅负责执行由 JobTracker 指派的任务。

但 Map/Reduce 作为一种离线计算框架，无法满足许多在线实时计算需求。目前在线计算主要基于两种模式研究大数据处理问题：一种基于关系型数据库，研究提高其扩展性，增加查询通量来满足大规模数据

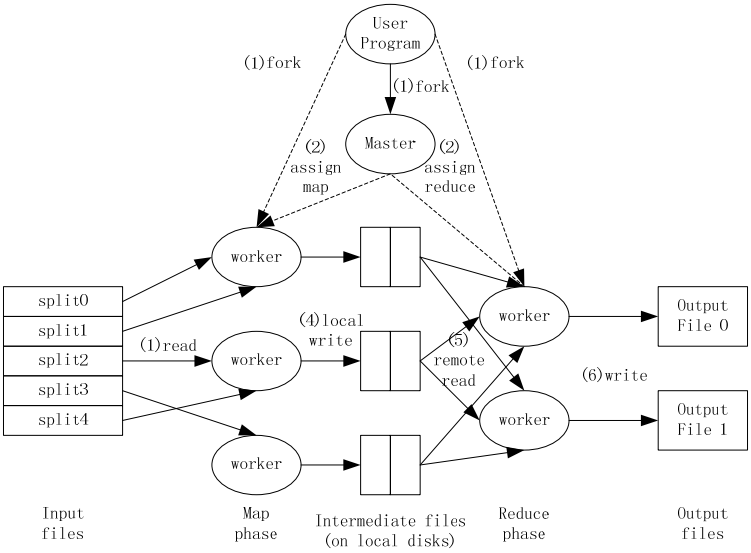


图 1 Map/Reduce 框架

处理需求；另一种基于新兴的 NoSQL 数据库，通过提高其查询能力、丰富查询功能来满足有大数据处理需求的应用。使用关系型数据库为底层存储引擎，上层对主键空间进行切片划分，数据库全局采用统一的哈希方式将请求分发到不同的存储节点已达到水平扩展的要求，这种方案一般不能对上层提供原存储引擎的全部查询能力。

另一方面，在处理复杂数据上，Map/Reduce 并不擅长，比如某些需要大量迭代运算的场景，在这些场景下，需要更适合的处理方式。

从 Map/Reduce 计算框架的优势和劣势可以看到，难以找到一种同时满足多种需求解决方案，解决问题往往需要多种不同处理方式的组合。因此，在指控系统中建立适合需求的数据处理平台，主要工作就是明确具体需求的应用场景、设计适合不同场景的解决方案以及对各种方案的灵活选择和组合。

### 3.3 数据分析

在指控系统智能化特征中，决策支持智能化是其重要组成。在数据规模增长迅猛的情况下，保证数据分析的效率和准确性愈发困难。在大数据环境下的数据分析，面临着两方面的问题，一是需要从更复杂的数据中来获取有价值的信息，二是分析的数据量更大，消耗的时间更长。为有效及时的数据分析提供支撑方案，让数据信息的价值得到充分体现和利用，对实现指控系统的决策支持智能化有着重要意义。

大数据中蕴含着丰富的价值，但是巨大的数量、数据的复杂性和模糊的分析目标都增大了任务的难度。如果可以通过一种简单的方式对数据规律进行直观展现，必将使大数据中的价值得到快速理解和发现，数据可视化就是这样的方式。可视化方法已经被证明为一种解决大规模数据分析的有效方法，并在实践中得到广泛的应用[6]。战场可视化将各种复杂的战场信息转化为容易理解的图形、符号提供给指挥员，提高决策效率；另一方面，通过可视化工具，作战命令能被迅速传递和理解，提高作战行动的主动性。

此外，当前学术界和商业界对于复杂数据智能分析技术和增量处理技术也关注较多。这里的复杂数据分析技术主要着眼于海量图数据的匹配分析和海量社交数据分析等内容，增量处理技术则面向准实时的搜索引擎查询需求，这些技术所分析处理的数据对象与指控系统中所分析的数据对象之间差异较大，难以直接应用于指控系统中。但这些技术的理论和设计思想，如增量处理技术的增量处理思想通过重点处理增删改的文档内容来提高频繁查询时的效率，可以考虑借鉴，为指控系统的数据分析提供理论基础和设计指导。

# 4 支撑指控智能化的数据处理平台构想

针对指控系统中的数据处理问题，本文结合大数据处理技术，提出以下支撑指控系统智能化的数据处理平台构想。

## 4.1 平台架构

整个数据处理平台的架构如图 2 所示。平台分为两部分：底层数据存储和前端数据处理。底层数据存储包括文件系统和数据库两个层次，在底层数据存储的设计中既包含了对已有关系型数据库（如 Oracle）的支持，也支持非关系型数据库（如 MongDB、Cassandra 等 NoSQL 数据库），这两类数据库建立在不同的文件系统之上，面向不同的应用需求。关系型数据库可用于存储读取频率不高的数据（如历史航迹数据等）、敏感数据（如口令、电子海图等关键数据），非关系型数据库则可用于支持具有较高实时性要求的数据（如实时传输的各类传感器数据）、更新频繁的数据（如战场态势信息）等。同时，在某些情况下还需要在不同类型的数据库之间进行数据桥接并实现集成。

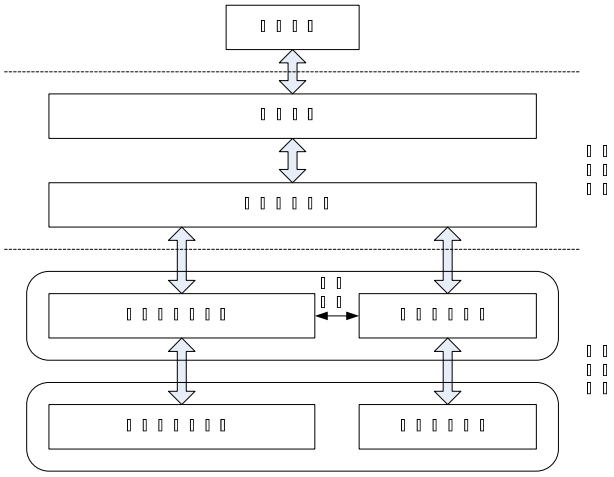

图 2 数据处理平台架构

前端数据处理包括数据处理框架和用户界面。数据处理框架主要支持对各种数据的预处理和搜索查询，用户界面为各种业务提供输入接口。前端数据处理所面向的业务主要为数据搜索与查询，支持实时、频繁的数据读写，可为数据分析提供满足实时性要求的大量数据预处理和查询支持。

## 4.2 构建原则

基于前述平台架构搭建具体平台时应遵循以下原则：

面向具体场景。实际应用中不同场景间的差异可能很大，具有通用性的解决方案或产品通常难以得到，应当根据不同的场景选择某种解决方案或将若干方案进行组合、折衷。

保证兼容性和易用性。兼容性是对已有基础设施及建立在之上的应用可继续发挥作用的保证，是对已有投入的最大限度利用；易用性是要尽可能地减少对用户使用的影响，简化用户操作复杂性、提升用户体验，同时也降低对用户进行重新培训的成本。

保证可扩展和高可靠性。数据规模的增长会一直持续，新的业务应用也会不断出现，因此处理平台需要具有良好的可扩展性，可以应对数据量的增加和新应用的加入。高可靠性是指控系统对于数据处理的必然要求，为了保证高可靠性，需要充分考虑各种容灾、抗毁要求，选择具有良好冗余设计的方案或实现。

# 5 总结

大数据处理技术作为当前已在商用领域得到成熟应用的数据处理技术，为应对指控系统在复杂海战场环境下的海量战场信息处理问题提供了有益的参考。通过对目前主流的大数据处理技术和产品进行梳理，可以看到由于商用需求与指控系统需求之间具有一些相同的基本点，在商用实践中使用的大数据处理技术是可以作为解决指控系统中数据处理问题的支撑技术的；但同时也是由于指控系统有着不同于商用需求之

处，现有商用大数据技术不能直接照搬到指控系统中使用，还需要进行改造或者仅仅只能借鉴其设计思想。

大数据的趋势是软硬件技术发展、业务需求增长等多方面因素共同推动的，大数据场景是今后指控系统所要面对的常态，对于大数据场景下的数据处理问题需要尽早研究并加以解决，从而提高指控系统的智能化程度，提升指控系统的作战效能。

**参考文献：**

[1]. 李德毅,林润华,李兵等. 云计算技术发展报告[M]. 北京:科学出版社, 2012:71-72.

[2]. 周晓方,陆嘉恒,李翠平等. 从数据管理视角看大数据挑战[J]. 中国计算机学会通讯, 2012, 8(9):16-20.

[3]. 马帅,李建欣,胡春明.大数据科学与工程的挑战与思考[J]. 中国计算机学会通讯, 2012, 8(9): 22-24.

[4]. 郭鹏. Cassandra 实战[M]. 北京:机械工业出版社, 2011:2-3.

[5]. 于元斌. 信息化条件下作战指挥效能研究[M]. 北京:军事科学出版社, 2010:21-21.

[6]. 袁晓如,张昕,肖何等. 可视化研究前沿及展望[J]. 科研信息化技术与应用, 2011, 2(4):3-13.: