Towards Data Intensive Scientific Computing in Defence and Security: A Review

Amandeep Kaur
TBRL-DRDO, Chandigarh, India

Mamta Mittal
G.B Pant Engineering College, New Delhi, India

P. K. Khosla
TBRL-DRDO, Chandigarh, India

Neerja Mittal
CSIR-CSIO, Chandigarh, India

ABSTRACT

Data intensive scientific computing is an emerging paradigm of science after experimental, theoretical and computational sciences. It refers to the extraction of knowledge from huge amount of data that is being generated with the proliferation of sensors and other computing devices in various areas like medical, engineering, military, defense and security, manufacturing, information technology, astronomy etc. This paper focuses on the various aspects of data intensive science in reference to defense and security. With the rise of smart warfare like drones equipped with missiles, smart bullets, robot soldiers and laser guns, the use of data acquisition, storage, processing, analysis, mining and visualization techniques have collectively envisioned the huge significance of data intensive science in this area. In this paper, the emergence of data intensive scientific computing as a new research paradigm and its significance in defense and security domain has been discussed along with some of the technological advances that account for the applicability of data intensive computing in this domain.

Keywords

Architectures, Big Data, Data Intensive Science, Visualization

1. INTRODUCTION

Data has become a driving force behind efficient decision making in various domains such as engineering, medical, astronomy, manufacturing, defence and security etc. To extract useful information from it, numerous acquisition, processing, analysis and visualization tools are available. This kind of computing has been termed as “Data Intensive Scientific Computing”, which recently emerged as the 4th paradigm of science after Experimental, Theoretical and Computational Sciences.

In defence sector, wide variety of instruments computational devices, drones, sensors etc. are being used for various applications like surveillance, reconnaissance, piloting, navigation, guidance, target identification and acquisition, weapon aiming, armament testing and evaluation etc. These instruments and devices could be mechanical, pneumatic, electrical, thermal and optical generating data with every use. This data could be operated upon by latest computing technologies and generated data could be preserved in an effective manner to carry out predictive analytics for efficient decision making.

However, the sensitivity of defence data requires a separate roadmap to implement such data-intensive computing techniques like Data analytics, Data Intensive Computing and Internet of Things. Standardized protocols and secure data storage architectures need to be developed in order to build the trust of defence organizations in utilization of new paradigms of computing technologies. Compartmentalization and balkanization of data according to its respective owners need to be ensured. In this paper, section 2 provides an overview of the emergence of data intensive scientific computing as a new paradigm, its major components, its significance in Defence sector and various computing frameworks. Section 3 discusses some data centric approaches for various defence applications and the concluding remarks have been given in section 4. The terms Data Intensive Science, Data Intensive Computing and Data Intensive Scientific computing have been used synonymously in this paper.

2. EMERGENCE OF DATA INTENSIVE SCIENTIFIC COMPUTING AS A RESEARCH PARADIGM

Data, now-a-days, is being envisioned as a new rival for oxygen, especially in the Information technology domain. But other domains like medical, astronomy, manufacturing, automotive and even defence are also being engulfed in this data deluge. According to research group IDG, 163 zettabytes of data will be created in a year by 2025(2). Thus the data intensive scientific computing is going to prove highly significant in solving a wide variety of issues and empower the decision making process. The major components of a Data Intensive Research process have been shown in Figure 1.

![Fig. 1: Components of Data Intensive Scientific Computing](image)

The acquisition of data is facilitated by various tools and methods such as sensor networks, satellite surveys, high throughput laboratory instruments, observation devices and...
supercomputers. The data so obtained is stored in the form of digital datasets, images, audio, videos and many other formats to be subjected to visualization and analysis.

The techniques of data mining, workflows, indexing, searching, graphics, analytics and visualization help in making sense of the collected data by determining the associations among the data and generating knowledge about the events for which those data was recorded. The knowledge so obtained is disseminated in one form or another to facilitate decision making as well resolve various problems and issues. The archival and preservation of data is equally important in a data intensive process as it helps in predictive analytics and also provides a base to build a powerful knowledge base[4].

2.1. Significance of Data Intensive Science in Defense and Security

Defense and Security sectors should also leverage the capabilities of data intensive science using various data management, analysis, visualization, new algorithms, existing softwares and tools in order to move from data to information to knowledge according to the knowledge pyramid.

An example of a shift towards data intensive science is of TNO, the Netherlands Organization for applied scientific research which has a world class onsite data acquisition facility for blast effects. An exclusive team has been deployed to setup and record highly dynamic events under any condition (indoor, outdoor, and offshore) [5]. Numerous analytical and processing approaches could be used on the acquired data to understand the dynamic events and other phenomenon captured through this facility. Many other such examples that quote the significance of data intensive computing in defence domain have been discussed below:-

- Armed forces of various countries are also investing in integrated Command, Control, Communications, Computers and Intelligence (C4I) capabilities which are meant to contribute towards the information overload which would require advanced processing capabilities [6].
- Large scale mapping requirements of military could be addressed through the exploitation of advanced technologies like Global positioning system (GPS), Geographic Information system (GIS) and Remote sensing[7].
- Accurate and updated Digital maps and imagery data could be a boon towards operationalization of fully automated digitized battlefield for defence forces.
- The military training methods are now moving towards E-learning platforms to enhance the physical preparation and productivity of defence forces by provisioning the visualization of battlefield.

The implementations of Army Clouds and Data Centers customized according to the security and privacy requirements of defence data have also envisioned a sharp rise in the amount of data being digitized per year. The data obtained from so many sources could be structured as well unstructured making it difficult to be processed by traditional data management systems.

Therefore, conventional algorithms and processing systems will not be able to address this data deluge known as Big Data. Integrated Military intelligence frameworks connected through secure communications networks need to be developed to strengthen the computational capabilities of defence research organizations as well as forces and security wings.

2.2. Data Intensive Computing Frameworks

Various dimensions of big data i.e. Volume, Velocity, Variety, Veracity and value have made it quite impossible for relational database management systems to process it. Therefore the computing techniques have become data centric in order to understand, evaluate, analyze, process and visualize various kinds of data efficiently and address the limitations of traditional systems. Numerous tools and frameworks have been developed for data intensive computing which have been divided into two major categories- Storage and Processing, some of which are listed in Table 1 [8].

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Category</th>
<th>Subcategory</th>
<th>Name of the package</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Storage</td>
<td>Distributed File System</td>
<td>Hadoop HDFS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NoSQL Databases</td>
<td>HBase and Cassandra</td>
</tr>
<tr>
<td>2</td>
<td>Processing</td>
<td>Batch Data</td>
<td>MapReduce and Spark</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Streaming Data</td>
<td>Spark Stream, Flink, Storm, S4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Graph Data</td>
<td>Giraph, GraphX, Flink, Graph Lab</td>
</tr>
</tbody>
</table>

The list of frameworks and processing tools is not limited to this table and is beyond the scope of this paper. Such tools and techniques need to be customized according to the Defence protocols and standards in coordination with private firms and technology vendors to keep in pace with the emerging trends and techniques. The adoption of latest data intensive computing methods and approaches will not only boost our defence capabilities but also speed up the processes and meet the time constraints.

3. DATA CENTRIC TECHNOLOGICAL ADVANCES IN DEFENCE SECTOR

The digitization and integration of various defence units to shift towards data-driven decision making has led to the development of numerous data centric approaches that could be quite useful in speeding up the activities and processes of defence sector. In this section, some important architectures and technological innovations have been discussed which have marked the beginning of data intensive computing in defence and security domain [18].

3.1. Graph Analytics Processor

This is a new processor architecture based on data intensive computing unveiled by Defense Advanced Research Projects Agency (DARPA) [10]. It is based on the idea that graph computing is an important tool to process connected data. Graph computing is a data analytics approach which is
3.5. Network Enabled Capability (NEC)

The objective of UK’s NEC[16] programme is to enhance military capability by the better exploitation of information. Such capabilities could also be put to use by various other defence organizations. This makes the geographically distant forces appear closer by sharing situational awareness, surveillance and other information. Body Area Networks is an important concept that adds to network enabled capability by providing vital information about the location, health and well being of soldiers deployed in extreme weather conditions[17].

3.6. Robust Encryption for Cyber Defense

With the expansion of datasets and sensitivity of the information contained in them, various cyber risks emerge. Considering the recent WannaCry[20] ransomware and the loss it incurred to a whole lot of people, processes and systems, there is a need of highly encrypted and secure architectures. A computer worm called Stuxnet[19] emerged in 2010 and had the capability to destroy industrial plants. Thus, frameworks involving point-to-point encryption, strict policy requirements and granular data access controls are now being put in place to counter the threats of data breaches and other vulnerabilities[21].

3.7. Multi sensor Data Fusion

Multi sensor data fusion techniques integrate data from multiple sensors and associated databases to achieve greater accuracies than single sensor systems. Various applications of data fusion in defense sector are automated target recognition for smart weaponry, guidance for autonomous vehicles, remote sensing, battlefield surveillance, and automated threat recognition systems[22,23].

Apart from these approaches, the breakthrough approaches of the Internet of Things (IoT) and Big Data Analytics are also depicting their prominence in building efficient defense systems and personnel. To leverage these technologies and approaches, standardization of secure data intensive and defense centric architectures need to be carried out. The involvement of private vendors and major technology giants is also important to keep up with the pace of technology.

4. CONCLUSIONS

Data intensive science has emerged as a highly significant computing paradigm which has transformed the decision making processes in various domains. However, defence and security sector is not leveraging its benefits to the most because of certain challenges such as, lack of standards and secure architectures to ensure data confidentiality, security and privacy. In this paper, the significance of data intensive scientific computing in defence sector has been discussed along with some associated computing frameworks. Recent data based technological advancements in defence and security have also been illustrated which need to be shaped into standard defence centric solutions.

5. REFERENCES


