Journal of Informatics and Web Engineering

Vol. 4 No. 2 (June 2025) eISSN: 2821-370X

Exploring Big Data Management Approaches and Applications: A Case Study of Real-Time Data Analytics in Air Traffic Management

Adeel Hashmi[1], Nouman Amjad[2], Muhammad Moiz Ullah Satti[3], Umar Hayat[4*], Anam Mumtaz[5]

[1]Department of Cloud Computing, Woodhouse Lane, Leeds, West Yorkshire, University of Leeds, UK

[2]Faculty of Network Security, Kaplan Business School, Melbourne, Australia

[3]Department of Computer Science, National University of Computer and Emerging Sciences (FAST-NUCES), Islamabad, Pakistan

[4,5]Department of Computer Science, Bahria University Islamabad, Pakistan

**corresponding author: (uhayat.buic@bahria.edu.pk; ORCiD: 0000-0002-1677-0144)*

Abstract - The rapid proliferation of digital devices has generated vast amounts of data, presenting significant challenges in collection, processing, and analysis that traditional systems struggle to overcome. This study investigates big data management approaches, explicitly focusing on technologies capable of efficiently handling real-time data at scale. Within the context of Air Operations, we propose a Hadoop-based architecture designed to support the Observe-Orient-Decide-Act (OODA) loop and enhance air traffic management. By leveraging a distributed system deployed on a cloud-based platform, we demonstrate a cost-effective solution for optimised data processing and improved decision-making capabilities. Our analysis highlights the advantages of using Hadoop's distributed file system (HDFS) for managing both structured and unstructured data generated by various sensors and devices. Additionally, we explore the integration of real-time processing technologies, such as Apache Kafka and Spark, to facilitate timely insights essential for operational effectiveness. Cloud deployment not only enhances resource accessibility but also offers flexibility and scalability, which are crucial for adapting to the dynamic nature of defence operations. We also address critical considerations for security and compliance when handling sensitive military data in cloud environments and recommend strategies to mitigate potential risks. The study concludes with recommendations for addressing future technological needs in big data management, including the incorporation of machine learning for predictive analytics and improved data visualisation tools. By implementing our proposed architecture, the military/ civil aviation can enhance its operational efficiency and decision-making processes, positioning itself to meet future challenges in an increasingly data-driven environment.

Keywords— Big Data, Hadoop, Data Analytics, Applications of Big Data, Big Data Processing, NoSQL, Horton works

Received: 1 February 2025; Accepted: 15 April 2025; Published: 16 June 2025

1. INTRODUCTION

In today's world, we are surrounded by the Internet of Things (IoTs). One cannot deny the massive increase in the use of social media in our daily lives. We are observing the viral growth of data that has undoubtedly reshaped the world.

A large number of people of various age groups, distinctly located around the world, use different platforms for exchanging and sharing views and news. The primary sources of data include the internet, social media, multispectral imagery, smart commercial and military gadgets, including advanced war machines, etc. Pakistan and international organisations must establish new capabilities and adapt to new technologies. The dramatic increase in the number and variety of information sources necessitates new strategies to support human role for the analysis of information and decision making.

Defence organisations with a large number of distinctly located people of various backgrounds use different platforms for exchanging views and sharing information. All such information is either stored in independent data banks or discarded without being brought to use. We need driveways to effectively capture, transform and utilise the information for planning and executing various analytical-based operations. Big data technology conceptually forms and drives the ecosystem of the modern world and will augment our ability to gather, structure, analyse and apply the processed data in our decision-making in the future.

Realizing the importance of information and the challenges to acquiring new capabilities, the Air Force must exploit digital information being created by mobile and wireless networks, social webs, private and public cloud services, sensor networks, and the IoT for effective and efficient decision-making. The Air force cannot neglect the critical linkage to the information collectors, and they will be instrumental in future strategies and operations. While the air force develops its operational capability, decision-makers must exploit the advantages of big data technologies for analytics-based operations. The exploitation of the information environment is inevitable for any future military operation.

2. BACKGROUND: KEY BIG DATA SYSTEM ELEMENTS

The journey of information and technology goes back to the Paleolithic tribe. Since then, humankind has brought evolutionary changes to technology. In 18,000 BC, Palaeolithic tribes used animal bones or wooden sticks to keep track of various activities. In 1663, John Grant performed the first statistical analysis of the plague and tried to find a way to make early discoveries about such outbreaks. Nikola Tesla, in 1926, came up with the idea of wireless technology and suggested an instrument that could be carried in pockets. Later, in 1928, Fritz Pfleumer invented a magnetic tape drive for storing data, which is still being used today. The concept of speech recognition was presented in 1962 by Shoebox Machine, which was used to interpret numbers and 16 English words. The USA planned to establish the world's first data centre, which was to be used to store user tax data along with fingerprints on tape drives. The emergence of relational databases came in 1970 when Edgar Codd presented a relational database for the storage of data in relational form.

In 1991, Tim Berners-Lee developed the concept of the World Wide Web (WWW), which brought a significant evolution in technology and mode of information flow [1]. A few years later, Michael Lesk pointed out that the WWW was growing 10-fold every year [2]. This viral growth of data gave birth to the big data idea in 1999. The term big data appeared for the first time in an Association for Computing Machinery (ACM) paper [3]. Web 2.0 was announced in 2005 [4]. In 2010, Eric Schmidt stated that the data being created every 48 hours was equal to the data that had been created until 2003 [5].

3. THE BIG DATA TECHNOLOGY AND THE V'S

The term big data can be defined in many ways; for example, according to the Oxford English Dictionary, it is the "Quantities, characters, or symbols on which operations are performed by a computer, considered collectively. Also (in non-technical contexts): information in digital form". The Merriam-Webster Dictionary defines it as a collection of "facts or information used usually to calculate, analyse, or plan something". Similarly, in Wikipedia, big data is defined as "An all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using on-hand data management tools or traditional data processing applications". The McKinsey big data survey explains big data as "Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse".

The phrase big data was first used by Doug Laney, who defined big data based on three main factors generally known as "Original Three Vs" [6]. i.e. volume, velocity and variety. Although the three V's have traditionally been used to define big data, as shown in Figure 1, more extra Vs have been proposed. However, these typically describe characteristics rather than being definitional. For example, IBM came up with the fourth V, i.e. the Veracity.

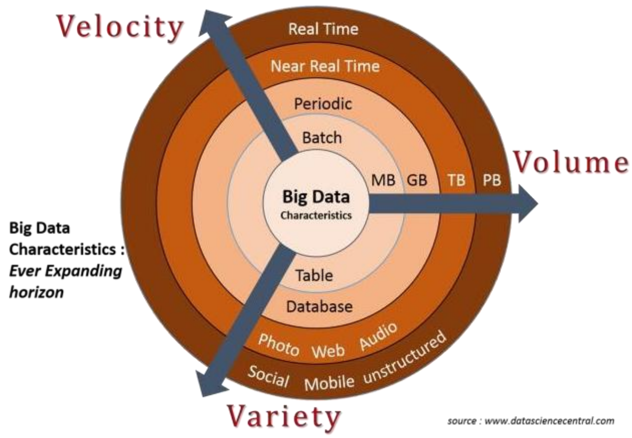

Figure 1. The Original 3Vs of Big Data [2].

3.1 Volume

There are many contributing factors to the surge in data. For example, transactions data stored in various silos over the years, Unstructured data streaming from social networks and the web, combined with relational data (comments, discussions, upvotes, etc.). According to research [7], 2.3 gigabytes of data is generated every 24 hours. Therefore, approximately 40 zettabytes of data will need storage by the end of 2020. Similarly, we have exponential growth of mobile users around the world. Close to 6 billion people use mobile phones, whereas the total population of the world is around 7 billion.

3.2 Velocity

The data is being generated and is acquired at an extraordinary speed. This data needs to be processed and analysed in real-time. The social networking software Twitter receives over 500 million tweets per day [8]. Data velocity is not only the speed at which data streams but also the time required to digest the data, which can maintain time-based relevance. High speed and velocity are about megabytes and gigabytes per second. Therefore, this data needs to be dealt with in a timely manner. However, the inconsistent data flows with periodic peaks make the job challenging to manage. According to New York Stock Exchange (NYSE) statistics, over 1TB of trading information is generated during every session. It is predicted that by 2020, 18.9 billion IoT devices will be generating data. Moreover, smart cars and aircraft have a large number of sensors that monitor the operational state and transmit it to various locations. All these sensors are sources of high-speed data.

3.3 Variety

The data is being generated in various types and formats from various IoT devices e.g. structured, semi-structured and unstructured. Structured data is the numeric data stored in traditional relational databases. Semi-structured e.g. XML or JSON is used to communicate between different platforms. Unstructured data is generated from sources like social media, images, video streams and audio, etc. According to research statistics [9], thirty billion contents (texts, images, videos, etc.) are shared by subscribers on Facebook every month. People tweet more than 400 million tweets per day. Similarly, the size of healthcare data has reached over 150 exabytes and it is predicted that by the end of 2020 there will be 420 million wearable, wireless health.

3.4 Veracity

The IoT devices generating big data can have unclear and ambiguous data with a lot of noise and garbage values. This can cause problems for data analysts. Therefore, strategies must be devised, and big data technologies must be used to make this data relevant and meaningful.

4. APPLICATIONS OF BIG DATA

In this section, we will present some effective applications of big data. There have been lots of research efforts to transform the social data into useful information [10]. Here the major concern, not only for the data analyst but also the general user is, that despite massive growth of social web data, there hasn't been significant efforts made in the management of this data available on the social networks. Mostly the contents available on the internet are either semi-structured or are totally unstructured. Thus, we are unable to make any logical inference from that. We cannot correlate pieces of information and extract more details out of the raw data that is floating around the networks. The semantic web is the solution to it, by making content in machine processable [11].

Twitter and Facebook are one of the most popular applications these days on the social web. There has been an exponential growth in the use of these applications. In Twitter, the number of users and tweets have increased many folds. According to recent research, on the average almost over one hundred and thirty thousand tweets are done in one single day [12]. The importance of semantics cannot be denied. As more and more data is being generated over the social web. Specifically, Twitter, creating huge unstructured information silos. The social media data will continue to grow.

The diverse and heterogeneous data will become more complicated and difficult to analyse. Due to the adoption of mobile advertising, biomedicines applications and smart cities the voice, images, videos and other unstructured data as part of this system. The 5V's of big data can easily be related to healthcare, wearable devices, retail, communication, smart grid, smart city etc [13].

The use of big data in today's world is imperative. The health care data is expected to double every coming month and half. The average health-related data generated by a person in lifetime will exceed one million gigabytes which is equivalent to 300 million books [14]. The shipments of sensor-based equipment related to health and fitness has increased from 107 million in 2012 to 515 million in 2017 [15]. The health care applications are presented in Figure 2.

Figure 2. Big Data Applications in Health Care [7].

Similarly, the use of wearable devices in the field of security, wellness, sports and fineness, lifestyle, computing, glamor and communication is increasing at the exponential rate e.g. smartwatch, HUD glasses, GPS shoes, heated clothing, Bluetooth jewellery, postures sensor belt, sun exposure monitor, identity verification, hotel key bands etc. are examples of above listed fields.

By the introduction of big data in retails, you can achieve traffic optimisation, supply chain management, brand optimisation, marketing, consumer co-presence and fleet management. Every mode of communication may it be car, ships, aircrafts, trains or supporting elements like toll collection units, travel assistance booths, passenger information centre, traffic signals etc. all will be generating data that needs to store, managed and analysed [16]. This will not be possible without big data technology stack.

The concept of smart cities as shown in Figure 3 has also evolved very rapidly starting from waste management, noise urban maps, smart lighting traffic congestion, smart parking water leakages, air pollution, fitness centres, intelligent shopping malls all will be part of interlinked networks. This network will generate data that needs to be analysed and processed.

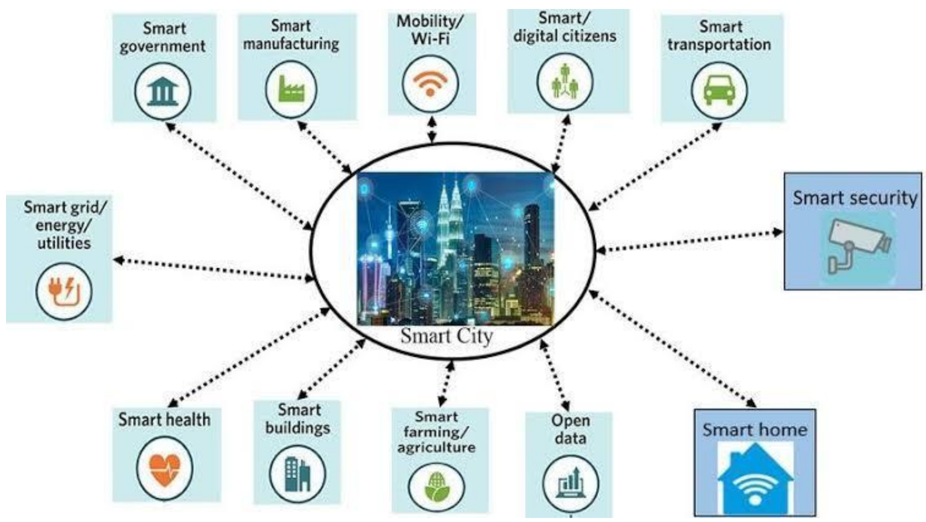

Figure 3. Smart City and Big Data [17].

5. TECHNOLOGY LANDSCAPE

The big data technology landscape can be categorised into infrastructure, analytics, applications, cross-infrastructural and open sources technologies [17], [18]. The most important part of this landscape is the data sources and API. In fact, it is the complete complex of various interdependent components that are used to store, process, visualize and present analysis.

Hadoop is an open-source framework based on Java, C and shell scripts. Hadoop was presented by Apache Software Foundation, and its first version was released in April 2006. Hadoop comprises various tools and frameworks that deal with data storing, processing, and analysing. The key features which make Hadoop one of the most powerful big data tools are open-source environment, highly scalable cluster, fault tolerance, high availability, cost effective, flexibility, faster data processing. The core components of Hadoop are HDFS, MapReduce, YARN, and Hadoop Common. All these tools work collectively to form a complete Hadoop ecosystem as shown in Table 1.

Apart from all of these, there are some other components like Ambari, Zookeeper and OoZie that makes Hadoop capable of processing large datasets as shown in Figure 4.

Table 1. Core Components of Hadoop

Component	Utilization
HDFS	It is used as a data storage layer. The data is stored as small data blocks. The HDFS splits the large datasets into small data-blocks with each data-block 128MB in size. Hadoop works on the MapReduce algorithm with the concept of master-slave architecture
MapReduce	It is the programming model that acts as a processing layer on Hadoop. It is designed for parallel data processing.
YARN	This layer YARN (Yet another Resource Negotiator) is responsible for management of resources and job scheduling. The performance of the data processing engine is dependent on the YARN framework.
PIG	It is a query-based language like SQL for structuring, processing and analysing large data sets. The concept of PIG technology was developed by Yahoo which is based on a pig Latin language.
HIVE	This is a query language for read-write operations of large data sets. HIVE can be used both for batch processing and real time data processing.
Apache HBase	It's a Hadoop Database (NoSQL database) that supports multivariant data

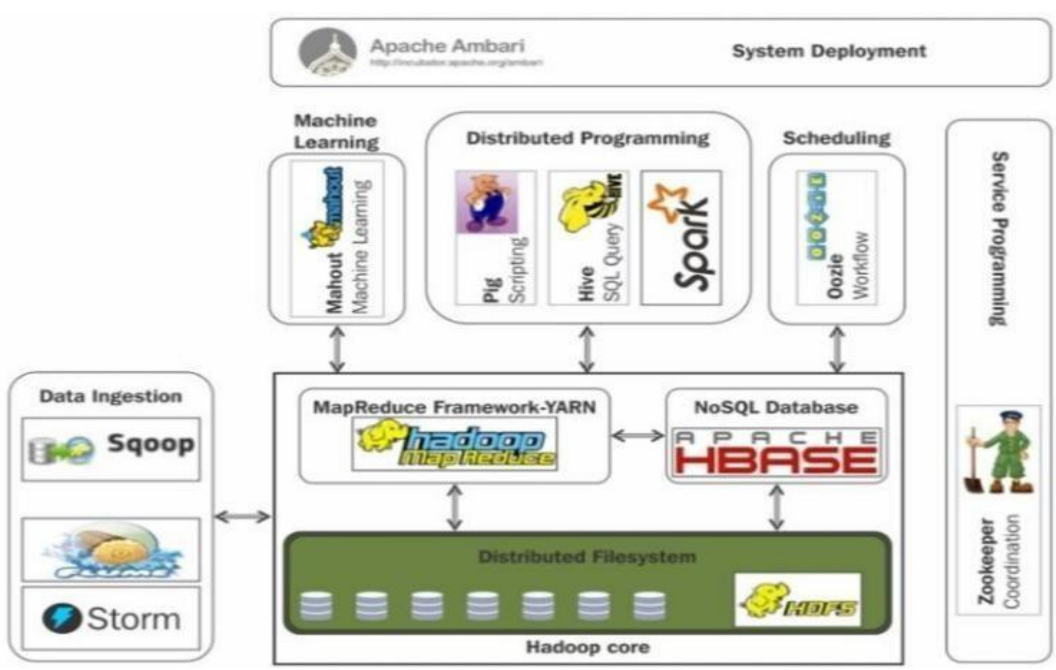

Figure 4. Hadoop Big Data Technology [7].

5.1 Hadoop Distributors

The major vendors distributing the Hadoop ecosystem includes Cloudera, Hortonworks, MapR and IBM. Cloudera was founded in 2008; it is listed as the leading company Hadoop software. Cloudera uses the open-source Hadoop software and makes proprietary changes related to management, security, and administration of the software. The largest competitor of Cloudera's in market presence. Hortonworks offers 100% open-source distribution of Hadoop, which gives users complete flexibility to customize some of its functionality. The third largest company in the market is MapR which is committed to offer Hadoop software with a balanced approach between high performance and scalability, while keeping ease of use as top priority.

5.2 Big Data Versus Traditional Data Warehouse

Big data and data warehouses are the main sources of data for analysis and intelligence reporting. Big data ecosystems can handle any form of raw data from multivariant sources, but data warehouses can only be used to store and manage processed data, as it must preserve the consistency of the data [19]. The main differences between the two are listed in Table 2.

Table 2. Data Warehouse versus Big Data

Data Warehouse	Big Data
Data warehouse is a data repository and generally used as a data store.	Big data is designed to handle huge data and prepare the repository.
Data warehouses can only handle structured data i.e. relational or not relational data.	Big data stack can manage multivariant data
The concept of a distributed file system doesn't apply in the case of the data warehouse.	Big data uses a distributed file system to load huge data in the form of blocks in a distributed way.

5.3 Big Data and Data Sciences

Data science is an evolution of statistics that deals with complex datasets using computer technologies [20], [21]. Big data deals with the extensive heterogeneous data coming from different sources that cannot be managed using standard database formats. Data science provides ways and approaches to digest and utilize the potential of big data, as presented in Figure 5.

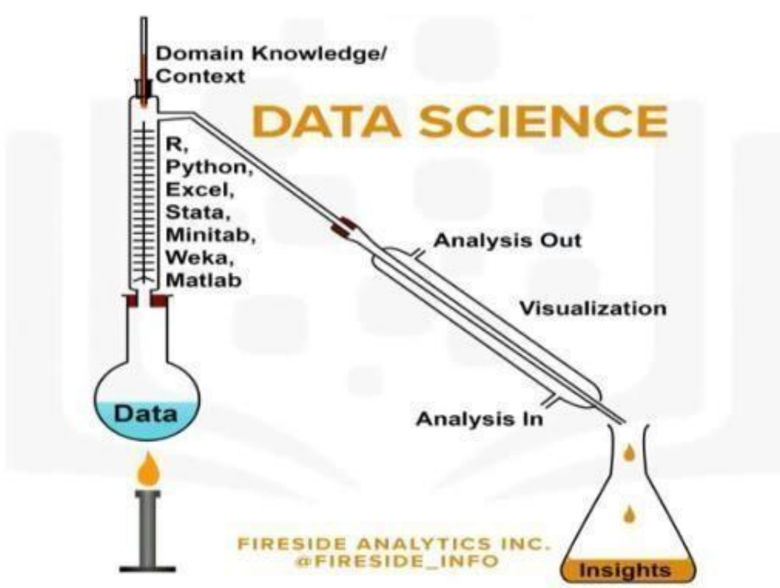

Figure 5. Data Science and Big Data Relationship [11].

Data science is used to extract value and meaningful information from data lakes. Big data provides the means and ways to manage large and complex data. However, digging out insight information from big data needs theoretical and experimental approaches of data sciences [22]. It makes use of machine learning and deep learning algorithms to get insights from big data. Both data science and big data address different concerns and hence must not be confused with each other.

5.4 Cloud Computing Support for Big Data

In this era of data explosion, various interrelated or independent mediums like social web, vlogs, blogs, air traffic systems, aircrafts, satellite imagery and other sensors are generating data every second. The data explosion caused by the various information collectors suggest that almost 90% of data present today have been generated in the last two years. This humongous amount of data will lead to storage problems and decision paralysis, where humans' ability to digest this amount of data will fail and we will not be able to manage and understand the data.

According to data scientists and statistics by DOMO [21], in 2020, the data being generated on the internet by various social web applications every minute is listed below that need a mechanism to store and process data. Approximately 2.5 million data sets are shared by Facebook users. Twitter manages nearly 300,000 tweets. Every 60 seconds over 72 hours of video content are uploaded on YouTube. Users receive or send more than 200 million email/text messages using various email platforms.

Cloud providers often use software as a service (SaaS) model to allow users to process data easily. The cloud-based infrastructure supports real-time processing of big data. It can take huge data sets from complex systems and interpret it in real-time. Another dependency between the two is that the power of the cloud allows big data to perform analytics very speedily. Cloud computing generally has powerful infrastructure that can be used for the application development by big data. Both these technologies play a very important role in our digital society.

5.5 Applicability of Big Data in Pakistan

The major organisations and companies are investing heavily in modern technologies like artificial intelligence, machine learning, data sciences, big data tools and predictive analysis that can augment decision making process and help in getting better solutions. The advent of these technologies improves organizational capacity to execute data intensive decisions.

According to the statistics, the financial companies spent over $6.4B in 2015 which is growing at the rate of 22% per annum. Similarly, the software/internet companies spent $2.8B in 2015 with annual growth of 26%. The communication/media invested $1.2B and the energy sector invested around $800 which has seen annual growth of 40% and 54% respectively [23].

The companies that are using big data and other analytical tools are likely to have two-times better financial performance, five-times better and faster decisions, three-times better execution of instructions and two-time better execution plans than those companies who are still thinking of adopting new technologies [24]. The major benefiting from big data includes eBay, Facebook, IBM, HP, The New York times, twitter, Amazon, Netflix, Intel, Yahoo, Samsung etc. Like many other countries, Pakistan is no exception to this problem. Different sectors of Pakistan are striving hard to gain a competitive edge by adopting this technology [25].

According to a statistical survey in 2018, there were 50.565 million bank accounts in Pakistan, where the total number of active ATMs and POS machines were 14,148 and 53,269 respectively. FBR performs has the responsibility to collect tax from all individuals and companies. This data needs to be assessed and analysed before applying tax laws. All this is impossible without big data technologies. Mobile telecommunications sector has seen large growth in Pakistan. The PTA is responsible to keep track and manage telecommunication information. Various mobile companies including Jazz and Telenor have over 59 million and 29.3 million subscribers respectively, while, Ufone has close to 23.1 million users on their network. All this customer data needs to be analysed to improve services and increase revenue of the company. No human work force can perform this without technological support.

NADRA is responsible for issuing the computerised national identity cards and maintaining sensitive information of the Pakistani citizens. It has one of the largest database infrastructures in Pakistan. Besides maintaining the huge record of individual data, NADRA also participate in major national activities e.g. In 2012, NADRA had removed approximately 37 million fake voters from the voting list and added approximately 36 million new adults to the voters list for the elections. Safe City Project Lahore, the administration of Lahore and Tec Access, one of the software companies in Pakistan had a joint venture to implement a big data solution for safe city Pakistan. The idea was to manage the data of vehicles, drivers and traffic moments. This will help in better traffic management and reduce traffic/ road crimes [26].

Last but not the least, our security and intelligence agencies are one of the most favourable organisations for big data technology. The war against terrorism has made the role of intelligence analysts very important. The monitoring and evaluating terrorist networks in Afghanistan and Iraq posed a distinct challenge. The various information collectives helped targeted operations against al-Qaeda and Taliban-linked groups.

5.6 Big Data in the Air Force

As we continue to move into the information age, the ability of nations to employ effective strategy to successfully deter and respond to threats has evolved. The Air Force must innovate and consider information as a distinct warfighting function. According to the Department of Defence USA, it is conceivable that information operations may surpass fire and manoeuvre in importance [27].

All the fighting forces specially air force cannot neglect the critical linkage to the information collectors as they will be instrumental in future strategies and operations. While the Air Force develops its operational capacity, higher command must strive hard for the induction of big data technologies and analytics to ensure future operations. The exploitation of the information environment is inevitable for any future military operation. Big data will play a vital role and will augment air force efforts to prepare and set the conditions for victory. Novel ISR task force model is presented in Figure 6.

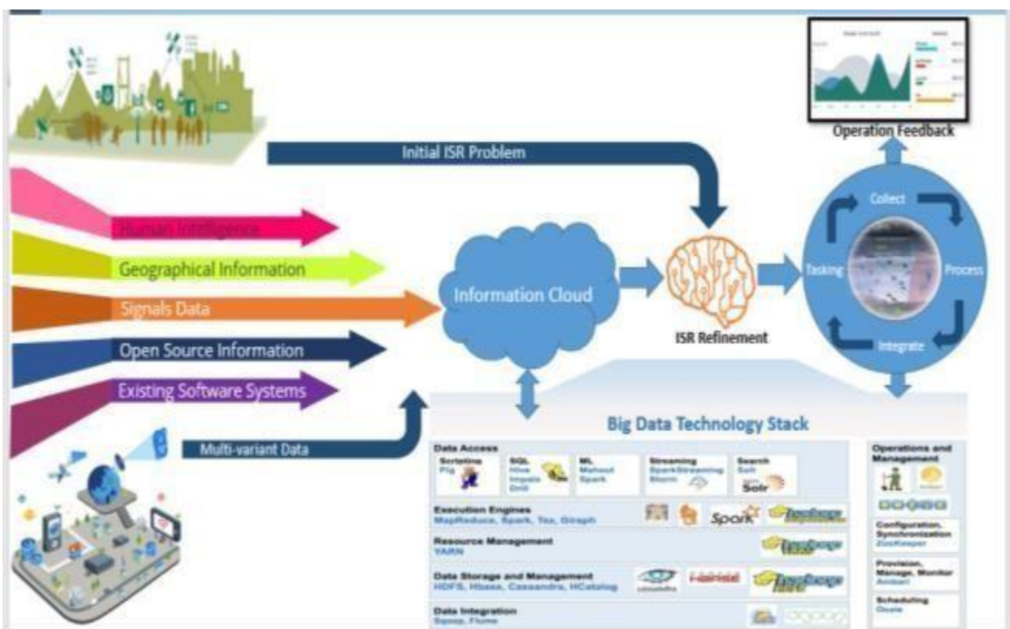

Figure 6. Novel ISR Task Force Model for Air Force

In the Air Force, generally an ISR is followed by an analysis of the data which may take days for operational-level analysis. This delay is against the principles of flexibility, agility, time and space. Further, various multivariant sources have grown more complex. Therefore, it is unlikely that traditional models for air tasking will remain effective and efficient in future. The new model for the air force, outlined in Figure 7, restructures the air operations planning process by incorporating big data analytics.

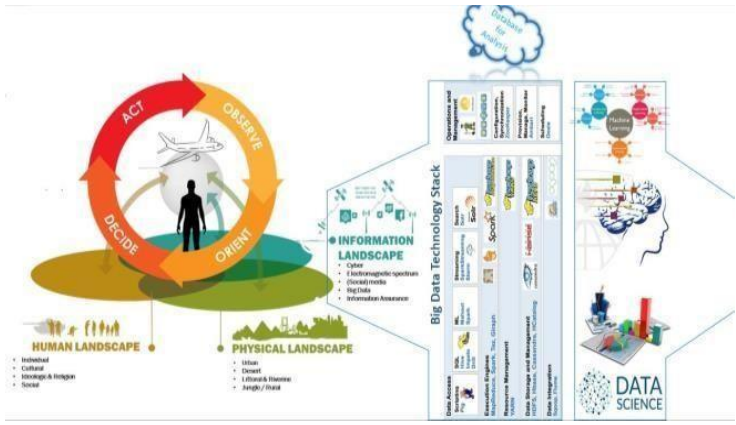

Figure 7. The Big data Framework for OODA Loop

In the novel framework of the OOD, a loop has been integrated with big data and analytics for effective decision-making. The multi-variant data will be verified and fused into the system using big data Hadoop architecture as presented in the figure. To make the information digestible and transform it into a holistic picture, the framework includes big data stream processing engine analytics and algorithms for real-time processing of the data. These technologies will empower us to detect and analyse patterns in real-time.

After the successful completion of observation and orientation phases, the decision step is all about identifying courses of action from gleaned data. We will apply analytics to the collected data to reveal potential patterns. For example, the analytics-based system will reveal that an air patrol pattern near the border line makes a U-turn at the same intersection every Monday morning. This unknown behavioural pattern can be identified and exploited by enemy aircrafts, putting our front line at risk. In the end, we must act. With an analytics-based OODA loop, we can act faster and respond to emerging threats more swiftly. By implementing big data technologies and data analytics into the operational decision cycle, we can elevate access to insight and intelligence.

Existing ICT systems and growing volume of data is gradually unfolding the challenge for its storage and processing. Huge data (unstructured and structured) in the form of text, documents, audio and video being generated from various platforms remains untapped. Taking this challenge as an opportunity, a four-node setup of this environment has been created using the open-source platform of Hortonworks. This environment can store multivariate data in a single storage pool thus making it ready for accumulative processing, predictive analytics and informed decision making. Till date three use cases have been implemented using the big data environment.

5.6.1 Use Case – I

In the first use case, a prototype of a security and surveillance system for vehicles was implemented. Using this implementation, car moment history, violations, driver details and locations of the vehicle can be traced. Moreover, this will help in identifying the friend and foe vehicles.

5.6.2 Use Case -II

In the second use case, personnel moment records were analysed. The temp duty data was visualized using the Tableau analytical tool. The dashboard presents a comprehensive report on temp duty records base wise and year wise. Moreover, the predictive analysis based on historical data suggests that, if the same trend of temp duty continues in the year 2021 the number of temp duties are likely to increase up to 10%.

5.6.3 Use Case-III

The third use case presents the trend analysis on the leave records of the personnel.

5.6.4 Under Study Use Case

The next use case under implementation is to predict component failure before the actual flight. Identify potential problems before they cause big problems and generate alerts for maintenance staff X days before predicted failure. Moreover, pinpointing failed items will lead to reduced maintenance time and reduce aircraft inoperative time drastically. The data sources that can help in smart decision-making include flying data, family history, pilot social networks data and medical history. Moreover, external factors, such as life, weather, and topographical information, will help predict flight performance. The established fact is that Information is ubiquitous, and information-based operations can badly affect an adversary and influence others actions. Certain assumptions were made like flight performance will be low due to long calls last night or less sleep, usually pilots with x medicine 7 days before don't give good performance in flights and pilots returned from holidays usually give good performance during flights/ pilots with family issues are prone to flying accidents. These are few hypothetical assumptions that will lead to many predictions eventually resulting in flight safety. The major organizations and companies are investing heavily in modern technologies like artificial intelligence, machine learning, data sciences, and big data.

6. ANALYSIS AND RECOMMENDATIONS

The decision making is not limited to the human and physical landscapes and information landscape must be involved in future decisions making cycle. The companies adopting the latest trends and technology will certainly gain huge competitive edges over their conventional rivalries. The use of smart devices and social networks have made big data imperative for our daily lifestyle. The same applies to defence organisations. To orchestrate military operations in all three dimensions, present command and control systems must be integrated with the information landscape for an effective operational cycle. Big data, private cloud, data sciences, deep learning and other analytical tools augment military decision-making at the strategic, tactical or operational levels [28]. By implementing big data technologies and data analytics into our operational decision cycle, the Air Force can put itself in an offensive position. This will help the air force to move towards the future with a clear vision where information explosion can be transformed to support the mission of the organisation. Considering the above stated facts and figures, flowing points are recommended: -

- Various organisations in Pakistan, especially defence/military organisations, will face decision paralysis due to the big data problems. By using the various big data solutions, we can help planners make more intelligent and more timely decisions. Big data is imperative and should be integrated with the Air Force decision support system to avoid decision paralysis.
- The big data, deep learning and data science setup, may be established with dedicated data scientists to identify intelligence gaps and find patterns in an enemy arrangement of battle.
- Predictive analytics empowers people to know more about the future and enables them to take pre-emptive measures in a calculated manner [29]. Therefore, predictive analytics must be used in the health sector to support and enable doctors to make pre-emptive measures in a calculated manner.
- Information analytics is imperative to improve aircraft maintenance cycle and flight safety [30]. Therefore, at all levels, information-based systems must be introduced to improve supply chain management.
- The human resources must be trained and educated on the importance of this technology. The capability development in innovative technologies must be done on a war ground basis.

7. CONCLUSION

The exponential increase in data and analytics from all sources of intelligence will make manual processes and the human workforce impractical and unnecessary. Nowadays, people have access to real-time information, and it is transforming our future and modern warfare. To conclude, the dramatic increase in the number and variety of information sources necessitates new strategies to support the human role in the analysis of information and decision-making. Big data will be imperative in future decision processes and will drive future actions.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Adeel Hashmi: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Nouman Amjad: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Muhammad Moiz: Project Administration, Writing – Review, Editing & Formatting;
Umar Hayat: Writing – Review;
Anam Mumtaz: Writing – Review.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] T. Berners-Lee, “Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor,” *HarperSanFrancisco*, 1999.

[2] M. Lesk, “How Much Information is There in the World?,” 1997. [Online]. Available: https://www.lesk.com/mlesk/ksg97/ksg.html.

[3] F. X. Diebold, “A Personal Perspective on the Origin (s) and Development of a Big Data: The Phenomenon, the Term, and the Discipline,” *Penn Institute for Economic Research, Department of Economics, University of Pennsylvania,* No. 13-003, 2012.

[4] T. O’Reilly, “What is Web 2.0? Design patterns and business models for the next generation of software,” *O'Reilly Media*, 2005. [Online]. Available: https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

[5] E. Schmidt, “Every two days we create as much information as we did up to 2003,” [Online]. Available: https://techonomy.com/conf/10/techonomy2010/.

[6] B. Marr, “Big data case study collection: 7 amazing companies that really get big data,” 2015

[7] A. De Mauro, M. Greco, and M. Grimaldi, “A formal definition of big data based on its essential features,” *Library review*, 65(3), pp.122-135. doi: 10.1108/LR-06-2015-0061.

[8] R. Holt, “Twitter in numbers,” *Telegraph*, 21 Mar. 2013. [Online]. Available: http://www.telegraph.co.uk/technology/twitter/9945505/Twitter-innumbers.html.

[9] B. Marr, “What is big data? A super simple explanation for everyone,” *Bernard Marr & Co.*, 2016.

[10] R. Roshini, and M. Raikar, “Map reduce based analysis of live website traffic integrated with improved performance for small files using Hadoop,”.

[11] A. Mumuni, and F. Mumuni, “Automated data processing and feature engineering for deep learning and big data applications: A survey,” *Journal of Information and Intelligence*, 2024. doi: 10.1016/j.jiixd.2024.01.002.

[12] J. Archenaa, and E.M. Anita, “A survey of big data analytics in healthcare and government,” *Procedia Computer Science*, vol. 50, pp. 408–413, 2015. doi: 10.1016/j.procs.2015.04.021.

[13] M. Manikandan, P. Venkatesh, T. Illakya, M. Krishnamoorthi, C. R. Senthilnathan, and K. Maran, “The significance of big data analytics in the global healthcare market,” in *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Apr. 2024, pp. 1–4. doi: 10.1109/IC3IoT60841.2024.10550417.

[14] Y. Zhang, “Application of big data in smart agriculture,” *Adv. Resour. Res.*, vol. 4, no. 2, pp. 221–230, 2024. doi: 10.50908/arr.4.2_221.

[15] K. Vijayalakshmi, S. Uma, R. Bhuvanya, and A. Suresh, “A demand for wearable devices in health care,” *Int. J. Eng. Technol*, 7(1-7), 4. 2018.

[16] K. Sravanthi and T.S. Reddy, “Applications of big data in various fields,” vol. 6, no. 5, pp. 4629–4632, 2015.

[17] C. Snijders, U. Matzat, and U. Reips, “Big Data: Big gaps of knowledge in the field of internet science,” *Int. J. Internet Sci.*, vol. 7, no. 1, pp. 1–5, 2012.

[18] S. Singh, T. Firdaus, and A. K. Sharma, “Survey on Big Data Using Data Mining,” *International Journal of Engineering Development and Research*, vol. 3, no. 4, pp. 135–143, 2015.

[19] M. R. Trifu, and M. L. Ivan, “Big data: Present and future,” *Database Systems Journal*, vol. 5, no. 1, pp. 32–41, May 2014.

[20] J. Y. Zhu, J. Xu, and V. O. Li, “A four-layer architecture for online and historical big data analytics,” In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pp. 634–639, 2016. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.115.

[21] G. Guerreiro, P. Figueiras, R. Silva, R. Costa, and R. Jardim-Goncalves, “An architecture for big data processing on intelligent transportation systems: An application scenario on highway traffic flows,” in *Proc. 2016 IEEE 8th Int. Conf. Intell. Syst. (IS)*, pp. 65–72, 2016. doi: 10.1109/IS.2016.7737393.

[22] T. C. Vance, T. Huang, and K. A. Butler, “Big data in Earth science: Emerging practice and promise,” *Science*, vol. 383, no. 6688, 2024. doi: 10.1126/science.adh9607.

[23] A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, “Big data technologies: A survey,” *Journal of King Saud University-Computer and Information Sciences*, 30(4), pp. 431-448, 2018. doi: 10.1016/j.jksuci.2017.06.001.

[24] S. Chaudhary, “Implementation and performance analysis of cloud,” *International Journal*, vol. 3, no. 9, pp. 317–322, 2013.

[25] L. Johansson, “Part 1: RabbitMQ for beginners – What is RabbitMQ?,” *CloudAMQP*, 2015. [Online]. Available: https://www.cloudamqp.com/blog/part1-rabbitmq-for-beginners-what-is-rabbitmq.html.

[26] L. U. *Laboshin, A. A. Lukashin, and V. S. Zaborovsky, “The big data approach to collecting and analyzing traffic data in large-scale networks,”* Procedia Computer Science, *vol. 103, pp. 536–542, 2017. doi: 10.1016/j.procs.2017.01.048.*

[27] *N. Brunetti-Lihach, “*Information Warfare: Past, Present, and Future*,” Nov. 14, 2018.*

[28] *G. Zeng, “Application of big data in intelligent traffic system,”* IOSR Journal of Computer Engineering*, vol. 17, no. 1, pp. 2278–661, 2015. doi: 10.9790/0661-17160104.*

[29] *R. Kumar, L. Adwani, S. Kumawat, and S. K. Jangir, “OpenNebula: Open source IaaS cloud computing software platforms,” In National Conference on Computational and Mathematical Sciences (COMPUTATIA-IV), Technically Sponsored By: ISITA and RAOPS, Jaipur, 2014. doi: 10.13140/2.1.3584.4805.*

[30] *M. L. Hamzeh Khazaei, S. Zareian, and R. Veleda, “Sipresk: A big data analytic platform for smart transportation,”* Procedia Computer Science*, vol. 166, Nov. 2016.*

BIOGRAPHIES OF AUTHORS

	**Adeel Ahmed Hashmi** has completed his MS Cloud Computing at the University of LEEDS, UK, with distinction. He has extensive hands-on experience in various technological stacks, such as the Cloud Solution Architect, Docker and containerisation, Big Data Solution Architect, Full Stack Data Scientist, DevOps, and HPC Solution Architect. In addition to his professional expertise, he has a strong interest in different sports and games. He has solidified his position as one of the best Golfers of Pakistan. He can be contacted at email: adeel.hashmi@gmail.com
	**Nouman Amjad** is a lecturer in Network Security at Kaplan Business School, specializing in network security, data science, and information systems. With a strong academic and research background, he is committed to preparing students for the dynamic field of cybersecurity. His research emphasizes strengthening security frameworks, developing data-driven threat detection methods, and enhancing information system resilience. Dedicated to merging theory with practice, he actively participates in research and industry collaborations. For academic or professional inquiries, reach out to him at itsnoumanamjad@yahoo.com.
	**Muhammad Moiz Ullah Satti** is a skilled Business Analyst with a strong foundation in SDLC, data analysis, and project management, dedicated to translating business objectives into effective technical solutions. Currently, he works at Giant Group, where he has hands-on experience in technical documentation (SRS, BRD, FRD). Holding a master's in data science and a bachelor's in computer engineering. He can be contacted at email: moizsatti41@gmail.com.
	**Umar Hayat** is a Machine Learning Engineer specializing in machine learning, computer vision, and application development. He holds a master's degree from Bahria University and currently leads ML projects at Synapbox company in Mexico. Previously, he worked as a Senior Data Scientist at the Pakistan Defence Organization, specializing in model development and data visualization. He has expertise in C#, Python, AWS, OpenCV, and TensorFlow and has received accolades, such as the Best Paper award at the 14th International Conference on Emerging Technology (2018). He can be contacted at email: uhayat.buic@bahria.edu.pk.
	**Anam Mumtaz** is a cybersecurity specialist with extensive experience in network security, firewall configurations, and 802.1x authentication. Currently serving as Lead Research Officer at the Pakistan Air Force, she leads the development of Network Management Systems (NMS) and configures Asterisk IPBX VoIP solutions. Previously, she worked on Cyber Threat Intelligence (CTI) and OSINT projects at NASTP and developed secure mobile applications at the Centre of Cryptology & Security. Anam is skilled in penetration testing, threat analysis, and technical writing. She is pursuing a master's in information security from Bahria University. She can be contacted at email: anammumtaz325@gmail.com.