Big Data: Related Technologies and Applications

Geetika Munjal and Manoj Kumar

1 Introduction

Data is described as "fundamental values or facts" taken from any person or agencies. Big Data is a type of data with an enormous volume. Big Data is the terminology used to express the immense volume and rapidly increasing data collection over time. Such data is so vast and complicated that no traditional data administration tool can save or process it efficiently. Big Data is hard to track. It represents the sum of all digital information, which has the inconvenience of storing, transporting and analysing. It is extensive and the technology is so tremendous that we have the challenge of creating today and the next generation's data storage tools and technologies.

In the epoch of Big Data [30], by analysing large quantities regarding data availability, it is possible to make rapid progress in many experimental methods and improve many organizations' efficiency and achievement. This creates unique possibilities for Big Data firms to gain more profound, sharper perspicacity that will empower decision-making, enhance the client's expertise and stimulate discovery. Firms are so surprised by the volume and type of data plus their activities that they strive to save data – evaluate, understand, and represent that in a significant way. The word "Big Data" is more than organized and agreement-aligned data. This involves videos, RFID records, communal schmoozing communications, demodulator channels, hunt indexes, natural happenings, pharmaceutical examines, "data

G. Munjal (✉)
Department of Computer Science, Amity University, Noida, Uttar Pradesh, India
e-mail: munjal.geetika@gmail.com; gmunjal@amity.edu

M. Kumar
School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India
e-mail: wss.manojkumar@gmail.com

F. Al-Turjman et al. (eds.), *Transforming Management with AI, Big-Data, and IoT*, https://doi.org/10.1007/978-3-030-86749-2_5

exhausts" – web surfers that track clicks through the Internet. Big Data technologies complement Business Intelligence (BI) instruments to open content from company's knowledge. It typically executes organized analytics and behind-the-scenes reflector of business administration, whereas Big Data analytics gives a forward-looking aspect, permitting corporations to forecast and perform on future events [1]. Big Data is a relative expression that illustrates a situation where the volume, velocity, and variety of data exceed the company's depository or its ability to make precise and appropriate decisions. Big Data, similar to Business Intelligence, Business Analytics, and Data Mining, has remodelled BI from reporting and determination support to forecast and next-move decision-making [2, 3].

Businesses use Big Data to chase gains, and authorities use it to serve the public welfare. Big Data provides the tools, techniques, technologies; IT structures to increase the exponential volume of diverse information and improve organizations' innovation and competitiveness in implementing sound and timely management decisions. Based on previous literature, it appears that factor models are the most common and widespread methods currently used for Big Data forecasting techniques such as the one seen in [28], whereas neural networks and Bayesian models are two other popular options.

The application of Big Data [31] can significantly benefit a small and medium-sized company, with businesses committed to the resources to execute Big Data technology. To make most of the Big Data, organizations need to develop their IT infrastructure to manage these new huge volumes, high accelerations, and various data origins and combine them with preceding enterprise data to analyse. Miscellaneous queries can be promptly resolved by applying Big Data and worldly wise analytics in a classified, memory, and lateral environment. The drift toward visualization-based data exploration tools can be realized by any business that wants to get the most value out of Big Data. Urban Big Data includes various types of datasets, such as air quality data, meteorological data, and weather forecast data. Raheja et al. [29] modelled the simulation of real-time air quality of any arbitrary location given environmental data and historical air quality data from very sparse monitoring locations.

Big Data is categorized into three elements: (a) data is diverse, (b) data cannot be classified as general relational databases, and (c) data is produced, compiled, and processed swiftly. Big Data is optimistic for business purposes and is immediately growing as a part of the IT industry. This has created considerable interest in various sectors such as healthcare mechanisms making, banking activities, social media, and satellite imaging. Traditionally, data was stored in an extremely organized arrangement to increase its information content. However, modern data quantities are driven by structured and semi-constructed data. Thus, the interchange between structured and unstructured data for analytics in relational systems of database management interrupts end-to-end processing.

The data collected at tremendous growth rate produces several key issues and challenges described, namely, faster data development, transfer velocity, assorted data, and security concerns. However, the advancement of data storage and tunnelling technologies enables the protection of this enhanced data. In aforementioned

conservation process, the characteristics of the data produced by the companies are altered. However, Big Data is yet at an early platform and has not been generally analysed.

2 Literature Review

In the information age, huge amounts of data are available for decision-makers. Big Data relates to datasets that are not only large but also big in size and variability, making them tricky to manage using conventional instruments and methods. With the speedy increase of data, there is a necessity to study and provide solutions to manage and gather value and information from these datasets.

This chapter surveys the most modern methods ripened for Big Data. Its mission is to assist you in choosing the correct consolidation of various technologies of Big Data and adapt them to the needs of their technology requirements and particular applications. Besides providing a global aspect of the major Big Data technologies, it also compares them to different system layers, such as Data Storage Layer, Data Processing Layer, Data Query Layer, Data Access Layer, and Management Layer.

This research's main intention is to examine the multiple Big Data technologies that can be applied to manage enormous volumes of data from various roots and improve the overall operation of the systems and its applications.

Big Data Technologies
Various Big Data technologies include:

(i) Column-based database: Traditional, queue-based databases are great for online trade processing at high modernized speeds, but as data volumes increase, they decrease in query performance, and data becomes more unorganized. Column-based databases stock data with a target on columns rather than rows, which allows for massive data squeezing and much faster inquiry times. The drawback of these databases is that they only permit batch updates with a much more delayed update time than traditional standards.

(ii) Schema-less database or NoSQL database: Several database types suit into this category, namely, the Value Store and Document Store, which focus on the depository and recovery of massive volumes of partly organized, or organized data. In NoSQL, this refers only to SQL, which covers a range of contrasting database technologies. The NoSQL database fields for processing relational ancestors, active, semi-structured data with minimal dormancy make them well suited to the Big Data ecosystem. NoSQL is generally described as operational and analytical. NoSQL is a custom function criterion for enhanced auxiliary functions based on an incomplete standard, where data can be processed at unrealistic times. Other big names in the NoSQL field are Cassandra, Oracle NoSQL and MongoDB.

(iii) Mass Parallel Processing (MPP) technologies process large volumes of data in parallel. A number of processors, each with their control system and memory,

work in contrasting parts of the identical program. MPP is a static process that requires a static database function between all the processors involved. During MPP, messages are exchanged between processors through the interconnection of data paths. For applications that allow multiple databases to be searched in parallel, an MPP system is considered superior to a symmetrically parallel system (SMP). There are examples of MPP architecture like supercomputers.

(iv) High-Performance Computing Cluster (HPCC) is a free source platform utilized for computing and rendering services to deal with Big Data workflows. The HPCC data design is determined by the customer end in line with the specifications. The HPCC system is projected and then planned to handle the most perplexing and data-intensive analytics-associated difficulties. The HPCC system is the only platform for a particular programming language for single architecture and data simulation. The HPCC system is planned to analyse large volumes of data to solve the perplexing problem of Big Data. The HPCC system is based on an organization control language with the Declarative and Procedural Nature of programming language.

(v) Hadoop is a free source software structure that processes massive amounts of data and processes large amounts of data. Hadoop presents the tools needed to develop and run applications. The data is divided into blocks and stored on multiple connected nodes that work together; this set-up is suggested as cluster.

The Hadoop cluster can traverse thousands of nodes. Calculations run across the cluster in parallel, which indicates that the task is split between nodes in the cluster. The Hadoop structure is penned in Java, which permits custom-written programs to locate system composed programs or a different language to refine data in order across millions of commodity servers. Hadoop employs a set of nodes to run the MapReduce programs in parallel. MapReduce program comprises two stages: the first map phase contains the processed input data, and the second reduction stage integrates the mediators in the result. To run MapReduce programs, each cluster node has a local CPU and a local file system [4, 9]. The data is divided into databases, collected in local files of different nodes, and evaluated for reliability. Local files create a file system termed as Hadoop Distributed File System (HDFS). Each cluster has several nodes ranging from thousands of machines to thousands of nodes. Hadoop can be combined with a fixed set of failover scenarios. Hadoop Ecosystem is a platform or suite that renders multiple assistance to resolve Big Data queries. This includes Apache designs and different commercial instruments and clarifications.

2.1 Components of Hadoop

The four main components of Hadoop are HDFS, MapReduce, Yarn and Hadoop Common in Fig. 1. Many tools or clarifications can be used to strengthen or promote these key components.

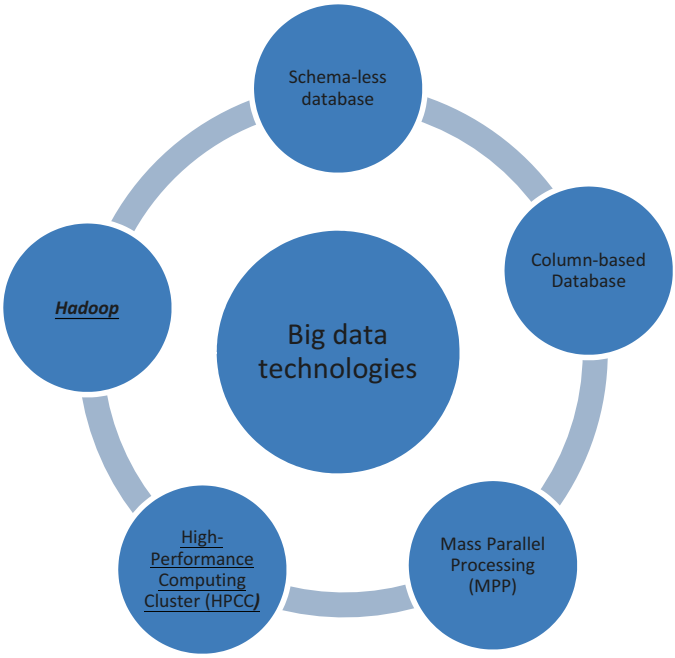

Fig. 1 Big data technologies

2.1.1 Management Layer

It helps in coordination and workflow, Zookeeper, Avro and Oozie as shown in Fig. 2. Zookeeper is a classified, public reference integration service for shared applications. It comprises of master and slave nodes and store composition data. Zookeeper promotes high efficiency and accessibility of data. It clarifies distribution programming and ensures a secure distribution area. The Zookeeper server runs in multiple groups. By its manageable port, Zookeeper facilitates quick, extensible, and secure high-performance computing framework coordination services for distributed systems [5]. For example, it offers a distributed set-up configuration management service, a naming service to find machines in massive clusters, a replication sync service to protect against loss of data and nodes, and a serialized access to resource-sharing locking service.

Avro is a time-worn method call and data serialization structure formed inside Apache's Hadoop design. From defining data types to organizing the data in a compressed binary composition, all activities are usually done by Avro. Data of Avro is saved in a register, with its plot collected so that any program can process it. Oozie: Organizing Apache Hadoop jobs is accomplished by Apache Oozie, which is a Java web application. Oozie couples diversified functions into one relevant task in a row. Oozie can schedule specific jobs for systems such as Java programs or shell scripts. It is an extensible, trustworthy, and extendable method [6].

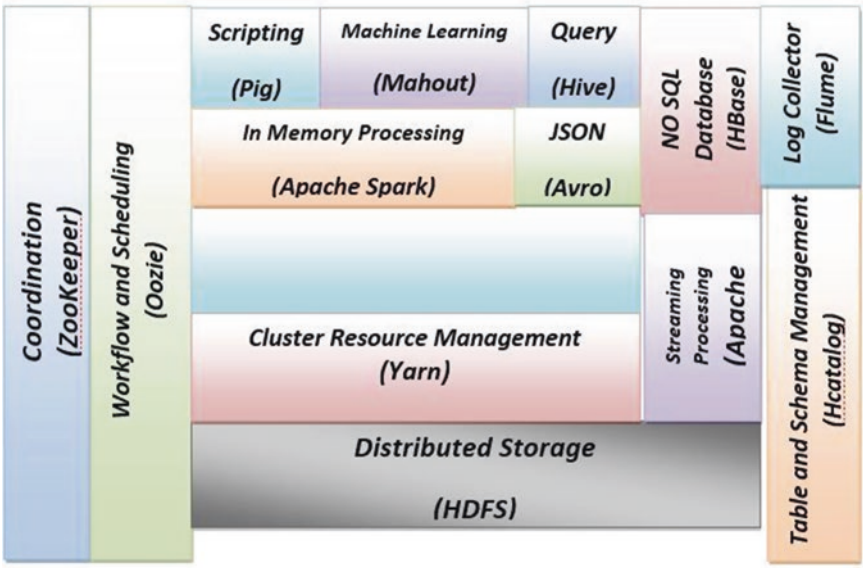

Fig. 2 Hadoop ecosystem

2.1.2 Data Storage Layer

HDFS (Hadoop Distributed File System) and HBase, where HBase is a dispersed, capable, scalable, NoSQL database that sits on the head of HFDS. It stores organized data in tables that contain lots of columns and rows. It can be used on Power Historical Discovery through huge datasets, exceptionally when the aspired data is in large amounts or inconsistent data. HBase is not a unique database and was not constructed for advertisement deals and additional real-time applications. It is available by the Java API and includes the ODBC and JDBC drivers [7]. When it comes to collecting massive volumes of data, we need more than one system, the main network system. Data can then be separated from multiple machines connected to each device over the network. This type of management is called a distributed file system to stock large amounts of data. Hadoop has its classified file system, called HDFS. This is the core of the Hadoop framework. It also eliminates excessive bandwidth across the cluster. It is a Java-based distributed file system that can stock all varieties of data without a previous company. There are high chances for the file system to be corrupted due to software bugs or human errors during software upgrades. The main purpose of creating snapshots in HDFS is to reduce the loss for data stored in the system during the upgrade.

2.1.3 Data Querying Layer

Pig, JAQL, and Hive; Apache Pig enables Apache Hadoop customers to draft complicated MapReduce changes practicing simply the expression of scripting termed as Pig Latin. Pig is an expensive program where the MapReduce structure is built, applied with the Hadoop platform. As in HLL by Pig, data accounts are examined, which is a top profile data processing method. The characteristic of the Pig is flaccidity, simply programmed, and self-optimizing. JAQL, the language announced on Hadoop, provides an inquiry language and promotes large-scale data processing. This transforms high-level inquiries into MapReduce tasks. It is intended to question semi-organized data according to the format of JSON. Such characteristics assure the processing of data, storage, translation, and transformation of data in JSON format. Facebook originally evolved hive. Organized data is processed by hive, a data depository foundation device. Hive is responsible to sum up Big Data and makes it simple to query and analyse. Hive generates its inquiry language called hiveQL. Hives are speedy, extensible, and compatible [8, 9].

2.1.4 Data Access Layer

It includes Data Ingestion; Chukwa is a framework for collecting data and analysing deal with MapReduce and HDFS. This frame is currently advancing beyond its developmental stage. Chukwa receives and processes data from distribution systems and reserves them in Hadoop. As an autonomous module, Chukwa is involved in the distribution of Apache Hadoop. It is set up on the upmost layer of the HDFS and MapReduce structure and acquires Hadoop's extensibility and hardihood. Flume is a disseminated, secure and accessible object to efficiently collect, integrate and move shells of log data. It has a convenient and adjustable architecture based on gushing data flow. This is a powerful dysfunction with tolerance and recovery mechanism with a tunable reliability mechanism [10]. It uses an asymptomatic, extensible data form that enables online application. When the run is off the line, it produces a run load file log. Whether this happens hundredth or millenary of times a day, huge amounts of log files can transmit data. The Flume tool can be stored for months or years of product runs for Apache Storm or similar day analysis in HDFS.

2.1.5 Data Streaming

It includes storm and spark where storm is a free source distribution system that holds the benefit of dealing with real-time data processing. The storm depends on the topology, which contains a network of spout, bolt, and streams. The bolt is utilized to process the input stream to create the output stream. Therefore, the storm is suitable for conversions on streams using "spouts" and "bolts".

The storm is simple to use, fast, extensible, and fault-tolerant system, and when more than one process fails, the storm instinctively relaunches. If the process crashes

frequently, it would be sent back to different machines and would be restarted again by storm. Real-time analysis, online machine learning, constant computing, and dispersed RPC are several such situations in which it can be practiced. To increase its performance, Spark is established on an in-memory system and is similar to Hadoop. It is a verified analytics platform that assures quick, simple to practice, and adaptable computing. Spark makes a complicated analysis of vast data collections through the In-Memory MapReduce system. The Spark Project includes task scheduling, memory administration, error retrieval, cooperation with storehouse systems, and more.

2.1.6 Data Processing Layer

It contains Hadoop MapReduce which provides the software infrastructure to make writing applications easier. This fault-tolerant trusted processor processes massive volumes of passive data in correspondence to a huge number of Ticklester Commodity hardware. The MapReduce job typically splits the input data into autonomous components, which are treated in parallel through map functions. Both input and output functions are saved in the file system. Scheduling tasks outline, supervise and perform indelible tasks [11, 12, 14]. The series MapReduce indicates less work is constantly done following the map job. The MapReduce framework's upper hand is cost-effectiveness, flexibility, and scalability due to its underlying parallel processing architecture.

2.1.7 Hadoop

YARN is an iconic part of the public source Hadoop platform for Big Data analytics accredited by the nonprofit Software Foundation. The essential components of Hadoop comprise the Central Library System, the Hadoop HDFS File Handling System, and Hadoop, MapReduce, which uses data to contain resources. Hadoop Yarn is defined as a grouping platform that manages assistant resources and scheduling assignments.

2.1.8 Mahout

Apache Mahout is a free source project that is mainly applied in the construction of extensible machine learning algorithms. It executes successful machine learning methods: advice, categorization, assembling. It is split into four central groups: grouping, refining, classification, and drilling of parallel periodic models. The Mahout Library refers to a sub-community that can be run in distribution mode and accomplished by MapReduce [13].

3 Applications of Big Data

Big Data is almost ubiquitous. Every business can implement Big Data analytics, such as health or normal living standards. Big Data is a track that can be applied in any field, and this huge amount of data can be used for one's benefit. The main applications of Big Data are posted beneath [15–17].

In *Agriculture*, various biotechnology companies uses sensor data [25, 28] to increase obtaining efficiencies. It collects plant tests and re-examines how plants respond to different changes in conditions. Its information is gathered around it for the quality and temperature of various plants, water level, soil system, growth, yield and quality of each plant in a proven ground. These recreations are approved to ensure the ideal environment for the correct quality sorts.

In Finance, relevant companies use external praise score when evaluating new acclaimed applications. Besides, banks are currently using their commendation score checks, which use a wide range of information for available customers and checks information of balances, charge cards, home loans, and corporations. Any monetary institution's performance depends on its data, and securing that data is one of the most challenging hurdles faced by any monetary institution. Data is the secondary commodity to them after wealth. Even ere Big Data became popular, the finance industry had previously conquered the tech sector. Additionally, financial institutions are amongst the lighthouse customers of Big Data and analytics. A big investment is at the heart of two of the most popular jargons in digital banking and payments. Big Data improves monetary institutions' core domains, such as deception discovery, uncertainty interpretation, algorithmic trading, and client fulfilment [18].

In Banking and Securities Industry, using customer data can also lead to privacy issues [27]. By looking at the incompatible parts between it, the exact bits of information, Big Data analytics uncover sensitive personal data. This shows that many financiers are wary of using Big Data because of isolation issues. Besides, the outsourcing of information research implementation or the sharing of customer communication and the product of a happy understanding opens up workplace security threats [26]. The Securities Exchange Commission (SEC) utilizes Big Data to observe the monetary market action. They are presently utilizing network analytics and universal language processors to capture unlawful commerce action in commercial markets. The industry depends on Big Data for exposure assessment: money lending, enterprise exposure administration, "know your customer" and demand reducing deceit. Big Data suppliers practicing exclusively in this industry involve Panopticon Software, Streambase Systems, Nice Optimize, and Quartet FS.

In Education, Big Data is the key to shape people's future and can remodel the education system for the betterment of the world. It is not just re-awakening collegiate skills, but also non-collegiate skills such as interpersonal abilities. Some best educational institutions are practicing Big Data as a means to reinvent their scholastic curriculum. Educational institutions can also trace student dropout rates and take the necessary steps to diminish this rate as much as feasible. In the state of the

differential application of Big Data in education, it can also be utilized to estimate professor potency to assure a delightful experience for learners and tutors. Professor performance can be accurately measured and marked upon numbers obtained by the student, the number of students, student goals, behaviour analysis, and countless additional factors. At the state level, the USA has been working to develop analytics to help students use the online curriculum to get the right curriculum. Among the Big Data suppliers in the industry are Newton and My Fit/Navion.

Manufacturing and natural resources: In the natural supplies industry, Big Data provides predictive modelling to be practiced to consolidate data and uniting massive amounts of geospatial data, graphical data, document, and transient data. Big Data has been adopted to address present production hurdles and achieve competitive benefits along with other advantages. Predictive manufacturing provides useless time near to zero and transparency. Huge amounts of data and sophisticated assessment tools are needed to be useful for the systematic process of data. The main advantages of practicing Big Data applications in the production industry are: i. Goods condition and bugs tracking ii. Stock preparation .iii. Production process and fault tracking iv. Yield anticipating v. Raising power productivity vi. Analysis and simulation of the latest production methods [21].

Government: In public services, Big Data has a broad variety of uses, including power research, economic market analysis, duplicity disclosure, health-associated analysis, and environmental security. Big Data is being practiced by the Social Security Administration (SSA) in the study of huge numbers of social disadvantage rights in the manner of structured data. Analytics can be utilized to process pharmaceutical knowledge faster and more accurately to make more agile decisions and identify unusual or counterfeit cases [22].

To recognize and analyse food-associated sicknesses and disease patterns, the Food and Drug Administration (FDA) uses Big Data. This enables it to respond quicker, leading to breakneck healing and less death of people around the world. Big Data is practiced for multiple cases by the Department of Homeland Security. Big Data from numerous regime companies are analysed and utilized to shield the homeland. Among the Big Data suppliers in the industry are Digital Reasoning, Socrates, and Hewlett-Packard.

Transportation industry, governments practice Big Data [19]: gridlock control, path outlining, smart transportation methods, bottleneck administration. Private-sector usage of Big Data in transportation: revenue administration, high-tech improvements, coordination, and competing for advantage. Personal usage of Big Data involves time-saving for energy savings and travel arrangements in tourism. Qualcomm and Manhattan Associates are among the industry's Big Data providers.

Energy and utility industry: Intelligent meter readers permit older meter readers to collect data every 15 minutes rather than once a day. This comminuted data is applied to better examine utilities, allowing more loyal user response and greater command over performance. In monopoly companies, Big Data application also enables better resource and labour pool administration, which can be used to identify shortcomings and correct them as quickly as feasible ere a complete breakdown

occurs. Alstom Siemens ABB and Cloudera are among Big Data suppliers in the industry.

Big Data Analytics has advanced healthcare by rendering personalized antibiotic and prescription analytics. Researchers are mining data to see which treatments are more beneficial for singular conditions, recognize drug aftereffects and other relevant knowledge that can assist patients and trim expenses. The amount of data along with mHealth, eHealth, and wearable technologies is growing at an accelerated increase rate. This comprises computerized health report data; patient produced data, detector data, and different types of data. By mapping healthcare data with geographic datasets, it is feasible to prognosticate disease progression in particular regions. According to estimates, it is easy to diagnose and prepare serums and vaccines. Some hospitals, such as Beth Israel, utilize data gathered from millions of patients from a cell phone app, to permit physicians to oppose multiple medical/laboratory tests for hospitalized patients to oppose evidence-based medicine. The University of Florida has used open public health data and Google Maps to build optical data that can help you quickly and effectively analyse healthcare information care used to detect the spread of chronic diseases. Hummedica, Explorer, and Cerner are among major data suppliers in the industry.

Media and entertainment: Several organizations in the media and entertainment industry are encountering innovative business models – creating, marketing, and distributing their content. This is due to the innovation of current users and the need to access content on any gadget, anytime, anywhere. Big Data presents facts regarding diverse people. Presently, publishing conditions are modifying ads and content to attract customers. This information is collected through several data tunnelling operations [22]. Media and entertainment industry is being benefited from Big Data utilization by foretelling what the audience wants, schedule optimization, growth of acquisition and retention and goal of advertising.

Internet of things (IoT) and Big Data: IoT is one of the major businesses for Big Data applications. Due to the immense diversity of objects, applications of IoT [29] are constantly unfolding. Nowadays, several Big Data applications are being supported by logistics companies. It is likely to trace the location of vehicles with sensors, wireless adapters, and GPS. Data collected from the IoT device provides a mapping of device interconnectivity. Several organizations and authorities have used it to improve the performance with the help of such mapping. IoT is also widely used as a tool for audiovisual data collection, and this audiovisual data is used in pharmaceutical and production contexts. Therefore, such data-driven applications permit organizations to monitor and supervise employees in addition to optimize delivery channels. This is done by tapping and joining numerous data, including prior driving practice [20].

Automobile: Big Data has taken full command of the automobile business and is operating it evenly. Big Data drives the automobile business astonishingly and has provided never before results. Big Data has supported the automobile business to accomplish things beyond our intelligence [20]. From examining trends to interpreting stock chain administration, looking after our customers, and realizing our

dreams of connected cars, Big Data is managing the automobile business well and truly insanely.

Telecom: Business is the heart of each digital innovation that is happening throughout the world. With the ever-gaining fame of smartphones, it has filled the telecom enterprise with huge volumes of data. And this data is like a gold mine, and telecom corporations require apprehension how to dig it. Through Big Data and Analytics, corporations can give clients seamless connectivity, thus eliminating all the network restrictions. Now with the guidance of Big Data and Analytics, companies can trace regions with low- and high-network traffic and, therefore, need to assure trouble-free network connectivity. Other enterprises, such as Big Data, have assisted the telecommunication industry to better comprehend its clients. The telecom industries are now offering customers as many customized offers as possible. Big Data is behind the data revolution we are currently experiencing [23, 24].

4 Conclusion and Future Scope

Big Data methods are invented to manage very huge and complicated datasets that cannot be processed by utilizing conventional systems. Originally, Big Data gained leverage to alter large volumes of automated data due to social media's bizarre extension. Since then, it has been utilized to process huge complicated datasets produced as a consequence of several experimental operations, construction methods, and network logs. In a very small period, Big Data has enrolled the technology platform and has established its presence in the field of technology and industry. Acknowledging the true potential of Big Data, companies are now passionately seeking to influence this technology to take advantage of the business and not face competitive risks in the long run. This chapter examines the notion of Big Data and the various technologies used to manage Big Data. For an industry that trades billions of dollars each year, Big Data is observed as a necessity rather than a luxury. It is no secret that Big Data is causing big changes in the business world. There are many advantages of Big Data, and it can be applied to regions that nobody thought of before.

Big Data is impacting the IT industry, just like some of the technologies that have been accomplished earlier. Extensive data produced from sensor-equipped devices, mobile phones, cloud computing, social media, and satellites can help various companies enhance their choice-making and drive their business to a different stage. Day after day data is produced so fast that traditional databases and other data storage systems gradually leave the storage, retrieve, and find relationships among data. The perplexity is that corporations require in-house skills and best methods. The downside of this is that Big Data has a service and advising boom.

The demand for solutions is so hot that all corporations are searching for a Big Data approach. Companies such as Google, Yahoo!, General Electric, Cornerstone, Microsoft, Kaggle, Facebook, Amazon are funding a lot in Big Data analysis and plans.

References

1. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
2. Liang Ting-Penga, & Liu Yu-Hsi. (2018). Research landscape of business intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications, 111*(30), 2–10.
3. Balakrishnan, S. (2019). An overview of agent based intelligent systems and its tools. *CSI Communications Magazine, 42*(10), 15–17.
4. Sun, Z., Sun, L., & Strang, K. (2016). Big Data analytics services for enhancing business intelligence. *The Journal of Computer Information Systems*, 1–8. https://doi.org/10.1080/08874417.2016.1220239
5. Normandeau, K. (2013). *Beyond volume. Variety and velocity is the issue of Big Data veracity*. Inside Big Data.
6. SAS Whitepaper. (2013). Big Data meets big data analytics. http://eric.univ-lyon2.fr/~ricco/cours/slides/sources/big-data-meets-big-data-analytics-105777.pdf
7. Oussousa, A., Benjelloun, F.-Z., Lahcen, A., & Belfkih, S. (2017/2018, June/October). Big Data technologies: A survey. *Journal of King Saud University – Computer and Information Sciences, 30*(4), 431–448.
8. Watson, H. J. (2014). Tutorial: Big Data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems, 34*, 1247–1268.
9. Storey, V. C., & Song, Y. (2017). Big Data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering, 108*, 50–67.
10. Ishwarappa, J. A. (2015). A brief introduction on Big Data 5Vs characteristics and hadoop technology. *Procedia Computer Science, 48*, 319–324. https://doi.org/10.1016/j.procs.2015.04.188
11. Yuri, D., Cees, D. L., & Peter, M. (2014). Defining architecture components of the Big Data ecosystem. In *Proceedings of 2014 international conference on collaboration technologies and systems (CTS)* (pp. 104–112).
12. Kumar, M., Punia, S., Thompson, S., Gopal, D., & Patan, R. (2020). Performance analysis of machine learning algorithms for Big Data classification. *International Journal of E-Health and Medical Communications (IJEHMC), 12*(4), 60–75.
13. Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with Big Data. *Proceedings of the VLDB Endowment, 5*(12), 2032–2033.
14. Punia, S. K., Kumar, M., & Sharma, A. (2021). Intelligent data analysis with classical machine learning. In S. S. Dash, S. Das, & B. K. Panigrahi (Eds.), *Intelligent computing and applications* (Advances in intelligent systems and computing) (Vol. 1172). Springer. https://doi.org/10.1007/978-981-15-5566-4_71
15. A personal perspective on the origin(s) and development of "Big Data": The phenomenon, the term, and the discipline (Scholarly paper no. ID 2202843) Social Science Research Network (2012).
16. Yadav, S. P., Mahato, D. P., & Linh, N. T. D. (2020). *Distributed artificial intelligence: A modern approach* (1st ed.). CRC Press. https://doi.org/10.1201/9781003038467
17. Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big Data, big analytics: Emerging business intelligence and analytic trends for today's businesses*. (Chinese edition 2014). Elsevier.
18. Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: Application of Big Data. *Journal of Big Data, 20*, 1–15.
19. Yadav, S. P. (2020). Vision-based detection, tracking and classification of vehicles. *IEIE Transactions on Smart Processing and Computing, 9*(6), 427–434, SCOPUS, ISSN: 2287-5255. https://doi.org/10.5573/IEIESPC.2020.9.6.427
20. Habib ur Rehman, M., Ibrar, Y., Salah, K., Imran, M., & Jayaraman, P. P. (2019). The role of Big Data analytics in industrial internet of things. *Future Generation Computer Systems, 99*, 247–259.

21. Perera, C., Liu, C. H., Jayawardena, S., & Chen, M. (2014). A survey on internet of things from industrial market perspective. *IEEE Access, 2*, 1660–1679.
22. Yadav, S. P., Agrawal, K. K., Bhati, B. S., et al. (2020). Blockchain-based cryptocurrency regulation: An overview. *Computational Economics*. https://doi.org/10.1007/s10614-020-10050-0
23. Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing Big Data strategies: A managerial perspective. *Business Horizons, 62*(3), 347–358.
24. Xu, L. D., & Duan, L. (2019). Big Data for cyber physical systems in industry 4.0: a survey. In *Enterprise information systems, 2019 – Enterprise information systems*. Taylor & Francis.
25. Cirillo, D., & Valencia, A. (2019). Big Data analytics for personalized medicine. *Current Opinion in Biotechnology, 58*, 161–167.
26. Bhardwaj, A., Al-Turjman, F., Kumar, M., Stephan, T., & Mostarda, L. (2020). Capturing-the-invisible (CTI): Behavior-based attacks recognition in IoT-oriented industrial control systems. *IEEE Access, 8*, 104956–104966. https://doi.org/10.1109/ACCESS.2020.2998983
27. Shankar, A., Pandiaraja, P., Sumathi, K., Stephan, T., & Sharma, P. (2020). Privacy preserving E-voting cloud system based on ID based encryption. *Peer-to-Peer Networking and Applications.* https://doi.org/10.1007/s12083-020-00977-4
28. Kumar, S., et al. (2021). Forecasting major impacts of COVID-19 pandemic on country-driven sectors: Challenges, lessons, and future roadmap. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-021-01530-7
29. Raheja, S., et al. (2021). Modeling and simulation of urban air quality with a 2-phase assessment technique. *Simulation Modelling Practice and Theory, 109*, 102281. https://doi.org/10.1016/j.simpat.2021.102281
30. Furht, B., & Villanustre, F. (2016). *Big Data technologies and applications*. Springer International Publishing. www.springer.com. https://www.springer.com/gp/book/9783319445489
31. Hung, P. C. K. (Ed.). (2016). *Big Data applications and use cases*. Springer International Publishing. www.springer.com. . https://www.springer.com/gp/book/9783319301440