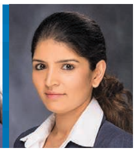
EDITORS
NORITA AHMAD American University of Sharjah, nahmad@aus.edu
PREETI CHAUHAN VP Technical Activities, IEEE Reliability Society, preeti.chauhan@ieee.org



NASA

Big Data: Present and Future

Preeti Chauhan, IEEE Senior Member
Mohit Sood, University of California, Berkeley

Big data in the 21st century will impact every individual, organization, and government. Organizations must invest in big data tools for business growth and efficiency while protecting data privacy as we continue toward digitization and datafication.

The term *big data* was first referenced in 1997 in an article by Michael Cox and David Ellsworth in the ACM digital library. The article discusses the challenges of visualization due to large data sets requiring high memory capacity. The authors referred to this as the problem of big data. Later in 2001, Doug Laney, an analyst with the Meta Group, published an article on data management with the "3Vs": volume, velocity, and variety; these terms went on to become the most commonly accepted definitions of big data. Over the years, variability and value were added as other key attributes of big data. In general terms, big data is a large complex set of data that requires additional computation to extract, analyze, and process to drive decision making.

Big data is broadly categorized into three data types: structured, unstructured, and semistructured. Structured data sets are made of clearly defined data types, which makes them easy to search and organize in relational databases. Some examples of structured data are phone numbers, street addresses, and Social Security numbers. Transactional data are another type of structured data and consist of, for example, sales orders, payments, returns, refunds, invoices, purchase orders, inventory-level changes, shipping documents, passport applications, credit card payments, and insurance claims.[1]

Unstructured data sets are those that are not easily searchable nor stored in structured database format. Such data could be textual, audio, or video and both human and machine generated. Human-generated unstructured data include audio and video data shared on YouTube, Instagram, Facebook, Twitter, and so on. Machine-generated unstructured data include satellite imagery, sensor, and digital surveillance data. Today, only 20% of the existing

Digital Object Identifier 10.1109/MC.2021.3057442
Date of current version: 9 April 2021

data are classified as unstructured, but at a 62% growth rate per year, by 2020, unstructured data will form 93% of the data sets.[2]

Lastly, semistructured data are a type that cannot be organized in relational databases, and they do not have a strict structural framework. However, they still have some structural properties. An example would be emails, which can be categorized based on, for instance, sender, subject, recipient, and send date and hence are structured. However, the content in emails falls under the unstructured data type, making emails a semistructured data set overall.

Big data is being created at an astounding pace to say the least. The total volume of data created, captured, copied, and consumed worldwide is expected to be 149 ZB in 2024, up almost two orders of magnitude from 2 ZB in 2010[3] (Figure 1). To put the size in perspective, it would take 181 million years if all of the existing data were to be downloaded. This explosion in data is a result of two macrotrends: an increase in Internet users and Internet-of-Things (IoT)-connected devices and a decrease in data storage and analysis costs, which came about due to a reduction in semiconductor computing (CPUs, GPUs, accelerators, and so on) and storage (memory) costs and advancements in network connectivity in the last two decades. Today, there are 4.7 billion Internet users[4] in the world, who are increasingly consuming and creating content on social media, search engines, online entertainment, and news. The Internet users generate about 6,123 TB of traffic every minute, which includes 185 million sent emails, 5.2 million Google searches, 305k Skype calls, and 84,000 photos uploaded on Instagram.[5] Apart from Internet usage, IoT-connected devices (sensors, smart cars, and so on) are expected to increase from 26.7 billion devices in 2020 to 75 billion devices in 2025.[6]

The cloud-computing industry has grown in lockstep with big data in the last two decades, with significant advances in data extraction, storage, and predictive and prescriptive analysis tools, including artificial intelligence (AI) and machine learning (ML). The global cloud-computing market size is expected to grow from US$371.4 billion in 2020 to US$832.1 billion by 2025, at a compound annual growth rate of 17.5%.[7] This growth rate is further expected to increase with accelerated cloud technology adoption by enterprises in sectors where the work-from-home initiative is helping to sustain enterprise business functions.

The global cloud-computing market size is expected to grow from US$371.4 billion in 2020 to US$832.1 billion by 2025, at a compound annual growth rate of 17.5%.

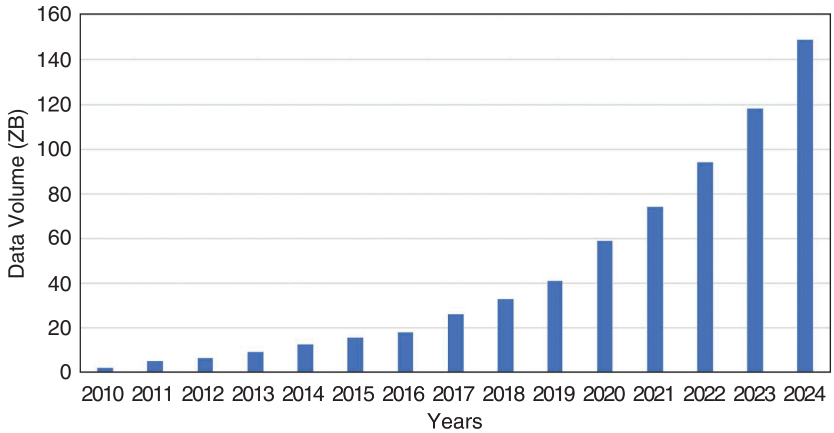


FIGURE 1. Big data volume growth from 2010 to 2024.[3] (Data source: Statistica.)

APPLICATIONS

Big data in today's world has tremendous potential to provide insights into almost all aspects of our lives, enabling smart decision making, cost reductions, future predictions, production-throughput improvements, and new product offerings. It ranges from providing personalized recommendations for best places to shop or eat based on user history, to playing a pivotal role for health agencies in managing the COVID-19 pandemic through contact tracing and hospital availability analysis. It is also widely accepted that companies focusing on big data analytics to create business values will succeed. This requires both strategic design and a well-thought-out architecture that can utilize the available data streams to meet specific business objectives, determine customer behavioral and usage patterns, and predict market trends.

Big data has enabled faster data processing and cost reduction by switching to cloud-based analytics, thus reducing the hardware and associated infrastructure needed for data storage and processing. Faster processing and internalization of complex data has enabled businesses to assess their competitors more quickly against their own offerings to make decisions in a continuously evolving business environment. The predictive capabilities offered by AI and ML have led companies to make future projections further out with higher accuracy and consistency.

The AI-powered user-pattern analysis from polls, surveys, Internet shopping, search history, location data, and so on has enabled the prediction of future human behavior and the provision of personalized recommendations for shopping, travel, restaurants, politics, weather, and even health. It is now possible to assess customer wants along with resulting satisfaction levels to deliver the right product. This has led many big companies to continuously innovate and launch customer-centric products on a consistent basis. Some of the industries benefiting from big data are health care, banking, media, retail, and energy. Other industries, such as medicine, construction, and transportation, are moving fast toward adopting and integrating big data analytics in their day-to-day operations and decision making.

The health-care industry benefits from big data by enabling the removal of redundant diagnoses from medical records, the early detection of diseases, and the ability to prevent virus outbreaks. The data have structured elements, like a patient's personal information and vitals, and unstructured elements, like X-ray and ultrasound records. Data are typically obtained from patient records or user generated from devices such as health apps,[8] Apple Watches, and Fitbits. The data are analyzed to help hospitals assess the effectiveness of therapies and drug administrations to improve future treatment plans. The most recent example is how health-care agencies and governments were able to do contact tracing for COVID-19 to follow the spread of the pandemic and help regulate social-distancing and shelter-in-place orders.

On the medical frontier, big data can be found at the leading edge of therapeutic and diagnostics research. For example, DeepMind, Google's deep-learning program, made a huge leap through its AlphaFold program to successfully determine the 3D shapes of proteins from their amino-acid sequence and solve a huge, decades-old challenge in biology. AlphaFold solved this problem after getting training on big data comprising approximately 170,000 protein structures.[9] Similarly, AI and big data platforms now provide the capability to sift through years of data to identify possible drugs that are already approved for treating certain diseases and help identify new molecules using this database to accelerate vaccine development.[10]

Although the banking sector has stringent data security regulations and has been relatively slow in adopting innovations, big data has started to play an important role in the banking business. The applications span fraud detection, customer behavior-pattern prediction, market trend detection, improved trade execution, and enhanced customer experience. Banks make use of both structured data (such as demographics, credit scores, and transaction types) and unstructured data (macrodevelopments, geopolitical news, and so on) to grow business and enhance client experience.

With the advent of online streaming, big data analytics has played a key role in driving the growth of the entertainment and media industry worldwide. Netflix, a popular streaming service, has experienced astronomical growth, with subscribers increasing from 21 million to 197 million globally in the last decade.[11] More recently, Disney's video-on-demand channel gained 74 million subscribers within a year of launch.[12] This growth has been due to both good content and big data. These platforms collect an incredible amount of data while their services are being used, to provide personalized recommendations for genres of shows and movies and improve customer engagement. Big data is also used to make decisions on which scripts or shows to produce or license by predicting viewership based on the content and performance of similar shows or movies in the past.[13]

Big data has also revolutionized the retail business forever. Online retail giants such as Amazon and Alibaba use big data to sift through millions of seller options in every product category to provide their users with enhanced experiences. Big data is also used in customer relationship management by almost 90% of businesses to enhance customer experience and increase sales.[14] Another example is Starbucks, which had harnessed big data from 30 million mobile app users and 20 million loyalty program members at the end of 2020.[15,16] The introduction of in-app payments, which provide valuable data about customer preferences, enables Starbucks to provide targeted offerings and rewards to its customers and increase overall sales.

TOOLS AND KEY PLAYERS

Organizations in general have to ingest both structured and unstructured data generated from disparate sources. Given this heterogeneity of big data, organizations need to make architectural choices about data storage and analytics solutions that provide both agility and flexibility.

On the medical frontier, big data can be found at the leading edge of therapeutic and diagnostics research.

The first architectural option is a data lake, which allows cost-effective storage of massive amounts of raw data that have an undefined or unclear purpose but are possibly needed for future use.[17] Data lakes offer fully redundant data storage infrastructure for storage and retrieval with accessibility across geographies, web spaces, or time horizons.[18] Data lakes have now become a prominent offering by cloud providers such as Microsoft (Azure Blob Storage), Amazon Web Services (AWS) (S3), Google (Cloud Storage), IBM (Cloud Object Storage), and Oracle (Cloud Infrastructure Object Storage) among others.

An alternative to a data lake is a not Structured Query Language (NoSQL)-based database that can handle nonstructural data with high availability and durability. A key benefit of NoSQL databases is horizontal scalability,[19] which allows seamless scaling of a single table over hundreds of servers and lower administrative overhead for operating and scaling distributed clusters.[18] While data lakes and databases have different advantages and are more suited to disparate business needs, companies such as Netflix use both to serve various requirements.[18] AWS DynamoDB, MongoDB, Google Cloud Bigtable, and Microsoft Azure Cosmos DB are among many NoSQL-based products that are currently available for database needs.

Once a data source is known, a data warehouse can be built using Extract, Transform, Load or Extract, Load, Transform operations.[20] A data warehouse consists of restructured data that are organized, easy to query, integrated from multiple sources, and of higher quality to ensure that robust reporting and data analysis can be performed for business intelligence (BI) or business analysis purposes. Data warehouse products are offered by all major cloud providers (AWS Redshift, Google Cloud's BigQuery, and Microsoft Azure SQL Data Warehouse), stand-alone providers such as Cloudera, and unique providers such as Snowflake that have the capability to integrate data from Amazon S3, Microsoft Azure, and Google Cloud platforms.[21]

The wide-ranging structured and unstructured stored big data is ultimately processed for predictive and prescriptive analysis to gain better understanding and insights in the application area.[22] The analysis spans from SQL-based BI reports, dashboards, and analysis to help business operations, to unstructured data processing using Apache Hadoop to solve data-intensive and computationally intensive problems[23] and AI and ML tools for building smart applications, such as predictive analytics, deep learning, image identification and classification, and natural language processing.

The rapid growth in the big data industry in the last decade has resulted in the availability of a multitude of tools to execute these analyses. While Tableau, AWS QuickSight, MicroStrategy Analytics, Microsoft Power BI, and Google Data Studio are among the most commonly used BI tools, Amazon EMR, Microsoft Azure HDInsight, and Cloudera Manager are some of the common unstructured data processing platforms that support big data frameworks, such as Apache Hadoop and Apache Spark. A wide-ranging suite of tools (Amazon's Lex, Polly, and Rekognition; Google's AutoML, AI Infrastructure, and Healthcare Natural Language,[24] and so on) are further available to conduct predictive analysis using AI and ML.

While it is arguable that such actions are user oriented within bounds of intended usage, there are grave risks around security breach and ethical usage of big data to influence people on deeply personal decisions.

TRENDS

Big data and data analytics will continue to grow in the coming years and are expected to be valued at US$230 billion by 2025.[25] An ever-growing number of organizations will continue to adopt and operationalize AI, resulting in a fivefold increase in streaming data and analytics infrastructure by 2024. The 2020 trends of growth in AI/ML, augmented analytics, edge-computing growth, in-memory computation, and continuous intelligence as well as growth or Spark and Databricks tools will continue in the next few years as well.

Augmented analytics utilizes AI and ML techniques to automate data preparation, sharing, and analytics, resulting in the transformation of complex, seemingly unusable data into smaller, usable data sets. It is estimated that augmented analytics markets will become a dominant driver for BI in 2021. Cloud-to-edge transition is also picking up to move away from centralized computing systems requiring high bandwidths. Edge computing will result in faster data analysis and reduced costs since it's better to extract and process data at the edge and then distribute them to relevant users/customers as needed. According to estimates,[26] by 2025, 75% of the enterprise data will be processed by edge computing. Due to the increasing need for real-time data analytics and the decreasing cost of memory, in-memory computing is expected to continue to grow in the coming years. This will be particularly helpful for business clients (banks, retailers, and utilities) for rapid identification of patterns.

Continuous intelligence is also expected to support automatic real-time data analysis and decision making via ML and continuous data analysis. It uses optimization, business rule management, event stream processing, and augmented analytics. It is predicted that more than 50% of new business systems will incorporate continuous intelligence by 2022.[27] Lastly, with the migration to cloud for data ingestion, analytics, and storage, traditional tools working on data center infrastructures such as Hadoop may no longer be the best option. Newer tools such as Spark, which can work with both data centers and cloud-based systems, will start to become more mainstream.

CHALLENGES

As big data is on a continuous-growth path, there are a few areas that need focus so that organizations and societies can continue to benefit their businesses, improve user experiences, and at the same time prevent privacy breaches, erroneous analyses, and disadvantages to small organizations trying to integrate big data into their day-to-day operations.

Ethical aspects of data collection, management, and application are continuously evolving and will influence industry practices in the next decade. Data privacy arising from big data is undoubtedly the biggest challenge affecting the current 4.7 billion Internet users. Big data is essentially the "big boss" in the online world, wherein our every action gets logged somewhere, often permanently, is analyzed, and influences our day-to-day decisions, often without any of us realizing it. The data include, but are not limited to, information about our favorite restaurants and cuisines, travel, shopping history, and search history.

Besides the personal information that users can choose to share, an individual's trails of disparate online data can be used not only to extract additional personal information but also draw inferences on how a person thinks by creating psychographic profiles. Using big data, the "big five" personality traits—openness, conscientiousness, extroversion, agreeableness, and neuroticism—can be determined with high accuracy. Researchers Kosinski and Wu created a model that predicts an average person's personality, sometimes even more accurately than their family and friends can. As an example, skin color (with 95% accuracy) and political affiliation (with 85% accuracy) could be predicted based on an average of 68 Facebook "likes" by a user.[28]

The combination of shared and deduced personal information on users is employed for highly personalized marketing to influence decisions on where to shop, eat, travel, and so forth. While it is arguable that such actions are user oriented within bounds of intended usage, there are grave risks around security breach and ethical usage of big data to influence people on deeply personal decisions. The biggest such situation in recent memory was the leak of the Facebook data of 87 million users—the largest known leak in Facebook history—by Cambridge Analytica to determine and sell psychological profiles of American voters for political campaigns.[29] These profiles, together with personal data, such as land registries, shopping data, and club memberships, purchased from data brokers like Acxiom and Experian,[28] helped to perform "behavioral microtargeting with psychographic messaging"[29] and influence voters. While the true effectiveness of such messaging has been a matter of public debate, the instance highlights how big data usage can significantly influence our lives on both a personal and a societal level.

Although data privacy affects 59% of the global population who are connected to the Internet,[30] only a very few, piecemeal policy responses to big data regulation, most prominently the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act, are in place to tackle this challenge. Data-collection practices and usage must be transparent, and companies must abide by them to ensure that user privacy and data breaches do not occur. The increased data mining from social platforms poses an increased risk of data misuse and loss or theft of sensitive personal information.

As companies venture into mainstream big data analytics, they will need to appropriately invest in the development of cybersecurity tools as well. GDPR was implemented in 2018 for data protection to regulate big data by empowering users to have the choice to decide with which businesses to share their data. The intent is to drive trustworthy data sharing with businesses, which in turn is expected to generate more reliable data and their associated analytics. At the same time, big data regulation should not come at the cost of efficiency and technological advancement and should be balanced to allow healthy and transparent sharing of data. A more universal policy framework for regulation of big data that balances the needs of organizations and the privacy of their customers should be deliberated and implemented by both national governments and international bodies.

A more universal policy framework for regulation of big data that balances the needs of organizations and the privacy of their customers should be deliberated and implemented by both national governments and international bodies.

Organizations currently working with big data or intending to delve into big data space encounter a myriad of challenges spanning adoption and operational issues due to the massive scale and analytics requirements, especially for unstructured data. While there have been continuous investments and advancements in utilizing structured and unstructured data, we are really at the tip of the iceberg for unstructured data. More unstructured data types are being added to the scope and are expected to dominate big data in the years to come, as we traverse the digital age dominated by social media and online platforms. There is a need for corporations to invest in the analytics for unstructured data to drive better BI. This requires big data management tools to have the right people, policies, and technologies to ensure the accuracy, security, and quality of data.

Organizations considering adopting big data often suffer from insufficient understanding and acceptance of it due to their legacy practices. It is therefore important to have a clear business use case and an expected value to be derived from big data analytics. This is especially relevant as the cost of big data analysis—whether done in house or outsourced—can be quite high as the data grow and expand. Organizations, especially nontechnological firms, also face considerable challenges making the right technology decisions themselves or even through external counsel from specialized firms.[31]

The other set of challenges can be attributed to the inherent massive scale of big data and its associated infrastructure needs. As organizations continuously generate astronomical amounts of data, the scalability of storage and analytic processes becomes increasingly difficult. It is anticipated that once data volumes surge to exabytes, data storage technologies and network bandwidths will become increasingly constrained and require continuous upgrades and technology advancements. Furthermore, the multifold aspects of big data management, including data distribution across geographies, access management, compliance, ownership, and governance, are becoming important from security, legal, and operational perspectives. Lastly, as organizations increasingly adopt big data, the heterogeneous expansion of data in both richness and variety will require the evolution of data architecture and analytics to process new data types. This increase in scale often compounds the noise, opacity, and relational nature of data, which increases the risk of data corruption and makes data validation, movement, and analysis very challenging.[32]

Big data analytics is another major problem often experienced when using unstructured data to arrive at desired results and conclusions. High data heterogeneity presents issues in determining a relevant data set for the analysis, preparing, or cleaning data and making sense of the unanticipated effects of outliers. Moreover, ascertaining the analytic method that is most appropriate for a problem or data set is not always clear. The decision-making process used to select from the wide range of analyses, such as predictive, prescriptive, and decision modeling; the different forms of analysis, such as quantitative, classification, visual; and the types of analysis, such as graph theory, social network analysis, behavioral analytics, econometric modeling, and control theory, among many others, can be both overwhelming and prone to error.[32]

Big data is undoubtedly one of the most defining trends of the 21st century, and it will impact every individual, organization, and government globally. The astronomical growth in the data generated from the Internet and IoT devices has provided huge opportunities to make improvements in decision making and efficiency in business operations and drive innovations in the research industry through analytic tools and AI/ML advancements. Big data is expected to be the next growth driver in all industries, from health and medicine to retail, banking, entertainment and media, and more. Along with the opportunities and growth associated with big data, organizations also face multiple challenges due to increased data size and complexity. This requires a continuous search for improved tools for data gathering, extraction, storage, and analytics and also necessitates organizations to be on the lookout for concerns of data breach and leaking of sensitive user information. Companies and organizations must address these challenges by continuing to invest in the development and adoption of big data tools and security and privacy practices to drive business improvements, protect user privacy, and ensure that they do not fall behind as the world marches onto digitization and datafication.

REFERENCES

1. D. McGilvray, *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. San Francisco, CA: Morgan Kaufmann, 2008.
2. "Structured vs. unstructured data—Best thing you need to know." ProWebScraper. https://prowebscraper.com/blog/structured-vs-unstructured-data-best-thing-you-need-to-know/ (accessed Jan. 30, 2021).
3. A. Holst. "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024." Statista. Dec. 3, 2020. https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed Jan. 30, 2021).
4. "Digital around the world." Datareportal. https://datareportal.com/global-digital-overview (accessed Jan. 30, 2021).
5. Internet live stats. https://www.internetlivestats.com (accessed Jan. 30, 2021).
6. "Top 10 big data trends of 2020." FinTech News. Nov. 2020. https://www.fintechnews.org/top-10-big-data-trends-of-2020/ (accessed Jan. 30, 2021).
7. "Cloud computing industry to grow from $371.4 billion in 2020 to $832.1 billion by 2025, at a CAGR of 17.5%." Globenewswire. Aug. 2020. https://www.globenewswire.com/news-release/2020/08/21/2081841/0/en/Cloud-Computing-Industry-to-Grow-from-371-4-Billion-in-2020-to-832-1-Billion-by-2025-at-a-CAGR-of-17-5.html (accessed Jan. 30, 2021).
8. "HealthKit." Apple. https://developer.apple.com/healthkit/ (accessed Jan. 30, 2021).

9. "AlphaFold: A solution to a 50-year-old grand challenge in biology." Deepmind. Nov. 2020. https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology (accessed Jan. 30, 2021.)
10. "Big data as a double-edged sword in the fight against COVID-19." ReadWrite. Apr. 2020, https://readwrite.com/2020/04/30/big-data-as-a-double-edged-sword-in-the-fight-against-covid-19/ (accessed Jan. 30, 2021).
11. J. Stoll, "Number of Netflix paid subscribers worldwide from 3rd quarter 2011 to 3rd quarter 2020." Statista. Jan. 2021. https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/ (accessed Jan. 30, 2021).
12. J. Bursztynsky. "Disney+ emerges as an early winner of streaming wars, expects up to 260 million subscribers by 2024." CNBC. Dec. 2020. https://www.cnbc.com/2020/12/11/after-showing-massive-growth-disney-hikes-5-year-subscriber-goal-.html (accessed Jan. 30, 2021).
13. E. Dans. "Netflix: Big data and playing a long game is proving a winning strategy." Forbes. Jan. 2020. https://www.forbes.com/sites/enriquedans/ (accessed Jan. 30, 2021).
14. D. Karr. "2020 CRM statistics: The uses, benefits & challenges of customer relationship management platforms." Aug. 2020. https://martech.zone/crm-statistics/ (accessed Jan. 30, 2021).
15. "Starbucks ramps up expansion efforts with focus on digital offerings." Business Insider. Dec. 2020. https://www.businessinsider.com/starbucks-drives-expansion-efforts-with-digital-offerings-2020-12 (accessed Jan. 30, 2021).
16. B. Pearson. "12 ways Starbucks' loyalty program has impacted the retail industry." Forbes. Dec. 2020. https://www.forbes.com/sites/bryanpearson/2020/12/16/12-holiday-gifts-from-the-starbucks-card/?sh=16f603df4534 (accessed Jan. 30, 2021).
17. "Data lake versus data warehouse." Talend. https://www.talend.com/resources/data-lake-vs-data-warehouse/#:~:text=Data%20lakes%20and%20data%20warehouses,processed%20for%20a%20specific%20purpose (accessed Jan. 30, 2021).
18. "Amazon DynamoDB vs Amazon S3." Stackshare. https://stackshare.io/stackups/amazon-dynamodb-vs-amazon-s3 (accessed Jan. 30, 2021).
19. "NoSQL vs relational databases." Mongodb. https://www.mongodb.com/scale/nosql-vs-relational-databases (accessed Jan. 30, 2021).
20. "Extract, transform, and load (ETL)." Microsoft. https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl (accessed Jan. 30, 2021).
21. "Snowflake: The data cloud." Snowflake. https://www.snowflake.com (accessed Jan. 30, 2020).
22. "Building a data lake on amazon web services (AWS)." AWS Cloud. https://pages.awscloud.com/rs/112-TZM-766/images/Building-a-data-lake-on-Amazon-Web-Services.pdf (accessed Jan. 30, 2021).
23. "What is Hadoop?" Amazon. https://aws.amazon.com/emr/details/hadoop/what-is-hadoop/ (accessed Jan. 30, 2021).
24. "AI and machine learning products." Google. https://cloud.google.com/products/ai#tab1 (accessed Jan. 30, 2021).
25. "Big data market worth $229.4 billion by 2025." Marketsand markets. https://www.marketsandmarkets.com/PressReleases/big-data.asp (accessed Jan. 30 2021).
26. "Edge computing by the numbers: 9 compelling stats." The enterprisers project. Apr. 2020, https://enterprisersproject.com/article/2020/4/edge-computing-9-compelling-stats (accessed Jan. 30, 2021).
27. "Gartner identifies Top 10 data and analytics technology trends for 2019." Gartner. https://www.gartner.com/en/newsroom/press-releases/2019-02-18-gartner-identifies-top-10-data-and-analytics-technolo (accessed Jan. 30, 2021).
28. H. Grassegger and M. Krogerus. "The data that turned the world upside down." Vice. Jan. 2017, https://www.vice.com/en/article/mg9vvn/how-our-likes-helped-trump-win (accessed Jan. 30, 2021).
29. N. Confessore. "Cambridge analytica and Facebook: The scandal and the fallout so far." NY Times. Apr. 2018. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html (accessed Jan. 30, 2021).
30. "Global digital population as of October 2020." Statista. https://www.statista.com/statistics/617136/digital-population-worldwide/#:~:text=Almost%204.66%20billion%20people%20were,percent%20of%20the%20global%20population (accessed Jan. 30, 2021).
31. A. Bekker. "The 'scary' seven: Big data challenges and ways to solve them." Mar. 2018, https://www.scnsoft.com/blog/big-data-challenges-and-their-solutions (accessed Jan. 30, 2021).
32. J. Alberto Espinosa, S Kaisler, F Armour, and W H. Money, "Big data redux: New issues and challenges moving forward," in *Proc. 52nd Hawaii Int. Conf. Syst. Sci.*, 2019 doi: 10.24251/HICSS.2019.131.

PREETI CHAUHAN is a Senior Member of IEEE, Santa Clara, California, USA. Contact her at preeti.chauhan@ieee.org.

MOHIT SOOD is at the University of California, Berkeley, Berkeley, California, 94720, USA. Contact him at mohit_sood@berkeley.edu.