|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Reference** | **Number of Citations (From Google Scholar)** | **Abstract** | Link |
| 1 | R. Cattell, Scalable SQL and NoSQL data stores, *ACM SIGMOD Record*, vol. 39, no. 4, pp. 12–27, 2011. | 25 | Computer clusters with the shared-nothing architecture are the major computing platforms for big data processing and analysis. In cluster computing, data partitioning and sampling are two fundamental strategies to speed up the computation of big data and increase scalability. In this paper, we present a comprehensive survey of the methods and techniques of data partitioning and sampling with respect to big data processing and analysis. We start with an overview of the mainstream big data frameworks on Hadoop clusters. The basic methods of data partitioning are then discussed including three classical horizontal partitioning schemes: range, hash, and random partitioning. | <https://ieeexplore.ieee.org/document/9007871> |
| 2 | F. Schneider and C. Weillner, "Big data and artificial intelligence", Nervenarzt, vol. 89, no. 8, pp. 859-860, 2018. | 15 | Application of big data and artificial intelligence has become one influence factor of English teaching, which have broken the balance of the teaching Eco-environment for English. In this article, the artificial intelligence and big data are introduced into English teaching to propose a new teaching Eco-environment construction method to meet the needs of the social development and international communication in English. In the proposed method, the characteristics of English teaching under big data environment are analyzed in detail. Then the big data technology is used to construct a new Eco-environment of English teaching to improve the teaching and learning quality. | <https://ieeexplore.ieee.org/document/9235580> |
| 3 | A. Alim, X. Zhao, J. Cho and F. Chen, "Uncertainty-aware opinion inference under adversarial attacks", IEEE BigData, pp. 6-15, 2019. | 15 | In the current technological era, huge amounts of big data are generated and collected from a wide variety of rich data sources. These big data can be of different levels of veracity in the sense that some of them are precise while some others are imprecise and uncertain. Embedded in these big data are useful information and valuable knowledge to be discovered. An example of these big data is healthcare and epidemiological data such as data related to patients who suffered from epidemic diseases like the coronavirus disease 2019 (COVID-19). Knowledge discovered from these epidemiological data-via data science techniques such as machine learning, data mining, and online analytical processing (OLAP)-helps researchers, epidemiologists and policy makers to get a better understanding of the disease, which may inspire them to come up ways to detect, control and combat the disease. | <https://ieeexplore.ieee.org/document/9378407> |
| 4 | Alamanda MS. Aspect-based sentiment analysis search engine for social media data. CSI Trans ICT. 2020;8(2):193–7. | 10 | There is an exponential growth in textual content generation every day in today's world. In-app messaging such as Telegram and WhatsApp, social media websites such as Instagram and Facebook, e-commerce websites like Amazon, Google searches, news publishing websites, and a variety of additional sources are the possible suppliers. Every instant, all these sources produce massive amounts of text data. The interpretation of such data can help business owners analyze the social outlook of their product, brand, or service and take necessary steps. The development of a consumer review summarization model using Natural Language Processing (NLP) techniques and Long short-term memory (LSTM) to present summarized data and help businesses obtain substantial insights into their consumers' behavior and choices is the topic of this research. A hybrid approach for analyzing sentiments is presented in this paper. | <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00680-6> |
| 5 | P. Buneman, S. Khanna and W.-C. Tan, "Why and where: A characterization of data provenance", Proc. Int. Conf. Database Theory (ICDT), pp. 316-330, 2001. | 13 | Nowadays, big data has become a hot research topic. It gives fresh impetus to the economic and social development. However, the huge value of big data also makes it the focus of attacks. Big data security incidents occur frequently in recent years. The security supervision capacities for big data do not match its important role. Data provenance which describes the origins of data and the process by which it arrived the current state, is an effective approach for data supervision. For the full use of provenance in big data supervision, a provenance model which defines the concepts used to represent the provenance types and relations is required to be built in advance, but current provenance models do not adapt to big data scenarios well. | <https://ieeexplore.ieee.org/document/9007438> |
| 6 | I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, The rise of ‘big data’ on cloud computing: Review and open research issues, *Inform. Syst.*, vol. 47, pp. 98–115, 2015 | 36 | With the recent advancements in computer technologies, the amount of data available is increasing day by day. However, excessive amounts of data create great challenges for users. Meanwhile, cloud computing services provide a powerful environment to store large volumes of data. They eliminate various requirements, such as dedicated space and maintenance of expensive computer hardware and software. Handling big data is a time-consuming task that requires large computational clusters to ensure successful data storage and processing. In this work, the definition, classification, and characteristics of big data are discussed, along with various cloud services, such as Microsoft Azure, Google Cloud, Amazon Web Services, International Business Machine cloud, Hortonworks, and MapR. A comparative analysis of various cloud-based big data frameworks is also performed. Various research challenges are defined in terms of distributed database storage, data security, heterogeneity, and data visualization. | <https://ieeexplore.ieee.org/document/9663258> |
| 7 | A. S. Alblawi and A. A. Alhamed, "Big data and learning analytics in higher education: Demystifying variety acquisition storage NLP and analytics", Proc. IEEE Conf. Big Data Analytics (ICBDA), pp. 124-129, Nov. 2017. | 14 | The proliferation of mobile devices and the rapid development of information and communication technologies (ICT) have seen increasingly large volume and variety of data being generated at an unprecedented pace. Big data have started to demonstrate significant values in higher education. This paper gives several contributions to the state-of-the-art for Big data in higher education and learning technologies research. Currently, there is no comprehensive survey or literature review for Big educational data. Most literature reviews from a few authors have focused on one of these fields: educational mining, learning analytics with discussions on one or two aspects such as Big data technologies without educational focus, social media data in education, etc. Most of these literature reviews are short and insufficient to provide more inclusive reviews for Big educational data. | <https://ieeexplore.ieee.org/document/9093868> |
| 8 | T. Devasia, V. T P and V. Hegde, "Prediction of students performance using educational data mining", Proc. Int. Conf. Data Mining Adv. Comput. (SAPIENCE), pp. 91-95, Mar. 2016. | 10 | Huge amounts of educational data are being produced, and a common challenge that many educational organizations confront, is finding an effective method to harness and analyze this data for continuously delivering enhanced education. Nowadays, the educational data is evolving and has become large in volume, wide in variety and high in velocity. This produced data needs to be handled in an efficient manner to extract value and make informed decisions. For that, this paper confronts such data as a big data challenge and presents a comprehensive platform tailored to perform educational big data analytical applications. Further, present an effective environment for non-data scientists and people in the educational sector to apply their demanding educational big data applications. | <https://ieeexplore.ieee.org/document/9393907> |
| 9 | Y. Zhang, Y. Xu, Z. Y. Dong, Z. Xu and K. P. Wong, "Intelligent early warning of power system dynamic insecurity risk: Toward optimal accuracy-earliness tradeoff", IEEE Trans. Ind. Informat., vol. 13, pp. 2544-2554, Oct. 2017. | 8 | In order to keep the bottom line of systemic financial risks and prevent the mitigation of major risks, this work focuses on the investigation of multi-source heterogeneous data fusion algorithms and cleaning technologies to establish a suitable style for data analysis and big data computation frame. In this paper, according to the above method, we provide the basis for early analysis of economic security. Utilizing the big data analysis, an emerging information technology method, we can be able to explore new risk early-warning methods, build a risk monitoring and early-warning platform and achieve scientific economic decision-making, so that the sources of economic risk in national economic security can be traced. | <https://ieeexplore.ieee.org/document/8967014> |
| 10 | R. Punnoose and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms", Int. J. Adv. Res. Artif. Intell., vol. 5, no. 9, pp. 5, 2016. | 10 | In the era of data science and big data analytics, people analytics help organizations and their human resources (HR) managers to reduce attrition by changing the way of attracting and retaining talent. In this context, employee attrition presents a critical problem and a big risk for organizations as it affects not only their productivity but also their planning continuity. In this context, the salient contributions of this research are as follows. Firstly, we propose a people analytics approach to predict employee attrition that shifts from a big data to a deep data context by focusing on data quality instead of its quantity. In fact, this deep data-driven approach is based on a mixed method to construct a relevant employee attrition model in order to identify key employee features influencing his/her attrition. | <https://ieeexplore.ieee.org/document/9409047> |
| 11 | Gennady Andrienko, Natalia Andrienko, Steven M. Drucker, Jean-Daniel Fekete, Danyel Fisher, et al.. Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications. BigVis 2020: Big Data Visual Exploration and Analytics, Mar 2020, Copenhagen, Denmark. ⟨hal-02568845⟩ | 34 | In the context of data visualization and analytics, this report outlines some of the challenges and emerging applications that arise  in the Big Data era. In particularly, fourteen distinguished scientists from academia and industry, and diverse related communities, i.e., Information Visualization, Human-Computer Interaction, Machine Learning, Data management & Mining, and Computer Graphics have been invited to express their opinions. | <https://dspace.mit.edu/bitstream/handle/1721.1/132286/BigVis2020_big_data_visualization_analytics_challenges_report.pdf?sequence=2&isAllowed=y> |
| 12 | J. Liu, T. Tang, W. Wang, B. Xu, X. Kong and F. Xia, "A Survey of Scholarly Data Visualization," in IEEE Access, vol. 6, pp. 19205-19221, 2018, doi: 10.1109/ACCESS.2018.2815030. | 68 | Scholarly information usually contains millions of raw data, such as authors, papers, citations, as well as scholarly networks. With the rapid growth of the digital publishing and harvesting, how to visually present the data efficiently becomes challenging. Nowadays, various visualization techniques can be easily applied on scholarly data visualization and visual analysis, which enables scientists to have a better way to represent the structure of scholarly data sets and reveal hidden patterns in the data. In this paper, we first introduce the basic concepts and the collection of scholarly data. Then, we provide a comprehensive overview of related data visualization tools, existing techniques, as well as systems for the analyzing volumes of diverse scholarly data. Finally, open issues are discussed to pursue new solutions for abundant and complicated scholarly data visualization, as well as techniques, that support a multitude of facets. | <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8314667> |
| 13 | U. Gupta and L. Fegaras, "Map-based graph analysis on MapReduce," 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA, 2013, pp. 24-30, doi: 10.1109/BigData.2013.6691550. | 16 | The MapReduce framework has become the de-facto framework for large-scale data analysis and data mining. One important area of data analysis is graph analysis. Many graphs of interest, such as the Web graph and Social Networks, are very large in size with millions of vertices and billions of edges. To cope with this vast amount of data, researchers have been using the MapReduce framework to analyse these graphs extensively. Unfortunately, most of these graph algorithms are iterative in nature, requiring repetitive MapReduce jobs. We introduce a new design pattern for a family of iterative graph algorithms for the MapReduce framework. Our method is to separate the immutable graph topology from the graph analysis results. Each MapReduce node participating in the graph analysis task reads the same graph partition at each iteration step, which is made local to the node, but it also reads all the current analysis results from the distributed file system (DFS). These results are correlated with the local graph partition using a merge-join and the new improved analysis results associated with only the nodes in the graph partition are generated and dumped to the DFS. Our algorithm requires one MapReduce job for pre-processing the graph and the repetition of one map-based MapReduce job for the actual analysis. | <https://csc.csudh.edu/btang/seminar/papers/BigD399.pdf> |
| 14 | Alexander Shkapsky, Mohan Yang, Matteo Interlandi, Hsuan Chiu, Tyson Condie, and Carlo Zaniolo. 2016. Big Data Analytics with Datalog Queries on Spark. In Proceedings of the 2016 International Conference on Management of Data (SIGMOD '16). Association for Computing Machinery, New York, NY, USA, 1135–1149. https://doi.org/10.1145/2882903.2915229 | 168 | There is great interest in exploiting the opportunity provided by cloud computing platforms for large-scale analytics. Among these platforms, Apache Spark is growing in popularity for machine learning and graph analytics. Developing efficient complex analytics in Spark requires deep understanding of both the algorithm at hand and the Spark API or subsystem APIs (e.g., Spark SQL, GraphX). Our BigDatalog system addresses the problem by providing concise declarative specification of complex queries amenable to efficient evaluation. Towards this goal, we propose compilation and optimization techniques that tackle the important problem of efficiently supporting recursion in Spark. We perform an experimental comparison with other state-of-the-art large-scale Datalog systems and verify the efficacy of our techniques and effectiveness of Spark in supporting Datalog-based analytics. | <https://dl.acm.org/doi/abs/10.1145/2882903.2915229> |
| 15 | Kouahla Z, Benrazek A-E, Ferrag MA, Farou B, Seridi H, Kurulay M, Anjum A, Asheralieva A. A Survey on Big IoT Data Indexing: Potential Solutions, Recent Advancements, and Open Issues. Future Internet. 2022; 14(1):19. https://doi.org/10.3390/fi14010019 | 4 | The past decade has been characterized by the growing volumes of data due to the widespread use of the Internet of Things (IoT) applications, which introduced many challenges for efficient data storage and management. Thus, the efficient indexing and searching of large data collections is a very topical and urgent issue. Such solutions can provide users with valuable information about IoT data. However, efficient retrieval and management of such information in terms of index size and search time require optimization of indexing schemes which is rather difficult to implement. The purpose of this paper is to examine and review existing indexing techniques for large-scale data. A taxonomy of indexing techniques is proposed to enable researchers to understand and select the techniques that will serve as a basis for designing a new indexing scheme. The real-world applications of the existing indexing techniques in different areas, such as health, business, scientific experiments, and social networks, are presented. Open problems and research challenges, e.g., privacy and large-scale data mining, are also discussed. | <https://www.mdpi.com/1999-5903/14/1/19?type=check_update&version=2> |
| 16 | Cumbane SP, Gidófalvi G. Review of Big Data and Processing Frameworks for Disaster Response Applications. ISPRS International Journal of Geo-Information. 2019; 8(9):387. https://doi.org/10.3390/ijgi8090387 | 25 | Natural hazards result in devastating losses in human life, environmental assets and personal, and regional and national economies. The availability of different big data such as satellite imageries, Global Positioning System (GPS) traces, mobile Call Detail Records (CDRs), social media posts, etc., in conjunction with advances in data analytic techniques (e.g., data mining and big data processing, machine learning and deep learning) can facilitate the extraction of geospatial information that is critical for rapid and effective disaster response. However, disaster response systems development usually requires the integration of data from different sources (streaming data sources and data sources at rest) with different characteristics and types, which consequently have different processing needs. Deciding which processing framework to use for a specific big data to perform a given task is usually a challenge for researchers from the disaster management field. Therefore, this paper contributes in four aspects. Firstly, potential big data sources are described and characterized. Secondly, the big data processing frameworks are characterized and grouped based on the sources of data they handle. Then, a short description of each big data processing framework is provided and a comparison of processing frameworks in each group is carried out considering the main aspects such as computing cluster architecture, data flow, data processing model, fault-tolerance, scalability, latency, back-pressure mechanism, programming languages, and support for machine learning libraries, which are related to specific processing needs. Finally, a link between big data and processing frameworks is established, based on the processing provisioning for essential tasks in the response phase of disaster management. | <https://www.mdpi.com/2220-9964/8/9/387> |
| 17 | Chris Stolte, Diane Tang, and Pat Hanrahan. 2002. Query, analysis, and visualization of hierarchically structured data using Polaris. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '02). Association for Computing Machinery, New York, NY, USA, 112–122. https://doi.org/10.1145/775047.775064 | 116 | In the last several years, large OLAP databases have become common in a variety of applications such as corporate data warehouses and scientific computing. To support interactive analysis, many of these databases are augmented with hierarchical structures that provide meaningful levels of abstraction that can be leveraged by both the computer and analyst. This hierarchical structure generates many challenges and opportunities in the design of systems for the query, analysis, and visualization of these databases.In this paper, we present an interactive visual exploration tool that facilitates exploratory analysis of data warehouses with rich hierarchical structure, such as might be stored in data cubes. We base this tool on Polaris, a system for rapidly constructing table-based graphical displays of multidimensional databases. Polaris builds visualizations using an algebraic formalism derived from the interface and interpreted as a set of queries to a database. We extend the user interface, algebraic formalism, and generation of data queries in Polaris to expose and take advantage of hierarchical structure. In the resulting system, analysts can navigate through the hierarchical projections of a database, rapidly and incrementally generating visualizations for each projection. | <https://graphics.stanford.edu/papers/polaris_olap/paper.pdf> |
| 18 | Antonino Galletta, Salma Allam, Lorenzo Carnevale, Moulay Ali Bekri, Rachid El Ouahbi, and Massimo Villari. 2018. An innovative methodology for big data visualization in oceanographic domain. In Proceedings of the International Conference on Geoinformatics and Data Analysis (ICGDA '18). Association for Computing Machinery, New York, NY, USA, 103–107. https://doi.org/10.1145/3220228.3220238 | 2 | Nowadays, thanks to new technologies, we are observing an explosion of data in different fields such as clinical, environmental and so on. In this context, a typical example of the well-known Big Data problem is represented by visualization. In this work, we propose an innovative platform for managing the oceanographic acquisitions. More specifically, we present two innovative visualization techniques: general overview and site specific observation. Experiments prove the goodness of the proposed system in terms both of performance and user experience. | <https://doi-org.proxy.library.kent.edu/10.1145/3220228.3220238> |
| 19 | Yanzhe Cheng, Fang Cherry Liu, Shan Jing, Weijia Xu, and Duen Horng Chau. 2018. Building Big Data Processing and Visualization Pipeline through Apache Zeppelin. In Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18). Association for Computing Machinery, New York, NY, USA, Article 57, 1–7. https://doi.org/10.1145/3219104.3229288 | 20 | Big data analytics pipeline becomes popular for large volume data processing, Apache Zeppelin provides an integrated environment for data ingestion, data discovery, data analytics and data visualization and collaboration with an extended framework which allows different programming languages and data processing back ends to be plugged in. The supported languages include Scala, Python, SQL, and Shell script as well as big data processing back ends including Hadoop, Spark and Hive. With the necessary tool sets, an interactive and dynamic data analysis can be done on the fly with heterogeneous programming interfaces. Although Zeppelin is great for code development and interactive analysis with small scale data set for proof-of-concept or use-case presentations, running the data processing pipeline in the batch mode is still needed for performance, robustness to fit in an automated workflow in some cases. We are developing a tool to convert Zeppelin notebook into a workflow with a set of codes that can run in a batch mode through command line interface without requiring running Zeppelin, so that the prototype code can be seamlessly deployed on the production cluster after demo stage. The entire workflow can be preserved, configured manually and run automatically. Zeppelin also provides a flexible way to integrate the visualization functionality, another contribution of this paper is to extend the Zeppelin's existing built-in visualization component for D3Network. With two added features described above, Zeppelin can help users to develop big data pipeline and visualizing graph data quickly and efficiently. | <https://doi-org.proxy.library.kent.edu/10.1145/3219104.3229288> |
| 20 | Z. Peng, Y. Chen, Z. Zhang, Q. Qiu and X. Han, "Implementation of Water Quality Management Platform for Aquaculture Based on Big Data," 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 2020, pp. 70-74, doi: 10.1109/CIBDA50819.2020.00024. | 4 | In order to ensure the quality and quantity of aquaculture, aquaculture farmers need to grasp the water quality in time. However, most farmers have to collect water quality data manually at present, and cannot store and reuse that information rapidly. This paper aims to use SpringBoot framework and JPA framework to build a big data platform of acquisition automation and visualization, which realizes the data analysis and display of heterogeneous water quality and breeding information. The platform can make the water quality prediction and real-time warning. Meanwhile, it realizes the management of robots, users and breeding experts. The application of this platform will bring better social benefits to aquaculture farmers. | <https://ieeexplore.ieee.org/document/9148352> |
| 21 | Xiaochuan Li, Spark Based Visualization analysis of Big intelligence Data, Shandong University, 2017. | 4 | With the rapid development of the Internet industry, more and more enterprises begin to realize the importance of data. Big data has gradually become an important reference for enterprises to understand their current situation and determine their future development direction. Big data visual statistical analysis platform refers to the system platform that completes most of the big data statistical analysis and shows the demand through the visual interface operation. The system platform mainly includes several functional modules such as visual ETL, visual construction site, authority management, data subscription and system monitoring. The research and development process is mainly based on the Spring framework. MySQL, Redis and HDFS are taken as data storage tools, and Apache Kylin, Spark and other data computing tools are used to carry out architecture design based on the core principles of high performance and high scalability. | <https://ieeexplore.ieee.org/document/9360991> |
| 22 | A. Gandomi and M. Haider, "Beyond the hype: Big data concepts methods and analytics", *Int. J. Inf. Manage.*, vol. 35, pp. 137-144, Apr. 2015. | 91 | Big data analytics (BDA) is a systematic approach for analyzing and identifying different patterns, relations, and trends within a large volume of data. In this paper, we apply BDA to criminal data where exploratory data analysis is conducted for visualization and trends prediction. Several the state-of-the-art data mining and deep learning techniques are used. Following statistical analysis and visualization, some interesting facts and patterns are discovered from criminal data in San Francisco, Chicago, and Philadelphia. The predictive results show that the Prophet model and Keras stateful LSTM perform better than neural network models, where the optimal size of the training data is found to be three years. These promising outcomes will benefit for police departments and law enforcement organizations to better understand crime issues and provide insights that will enable them to track activities, predict the likelihood of incidents, effectively deploy resources and optimize the decision making process. | <https://ieeexplore.ieee.org/document/8768367> |
| 23 | Zhihan Lv, Houbing Song, Basanta-Val, Anthony Steed and Minho Jo, "Next-Generation Big Data Analytics: State of the Art Challenges and Future Research Topics", *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1891-1899, 2017. | 2 | Exponential growth rate in the data generation in diverse fields revolutionized the way the analytics tools and machine learning algorithms applied in the Big Data. Considering the pace at which data is generated and the variety of data, identifying the right data at the right time and the relationship is crucial to make decisions. Identifying the data type and the relationship between the parameters is challenging as the size of the real time data is massive and dynamic in nature. Data Visualization as part of Big Data Exploratory analysis helps to identify the relationship and to understand the characteristics of big data in an effective way. In other words, data visualization is the graphical representation of data, it links the data availability and data analysis, organizes and presents important findings from the data. As plenty of visualization packages and tools are available, choosing the right tool is very important. gglot2 is one among the statistical graphics tool for data visualization as the working model is entirely based on grammar of graphics. The unique feature of ggplot is the layered approach, as the each component is highly interdependent on the other and so it helps in the step by step analysis of the data. With this feature available ggplot2 was used in the Real time Public Distribution System data to identify the relationship between various parameters and to understand the behavior pattern. | <https://ieeexplore.ieee.org/document/9065491> |
| 24 | T. Shafer, The 42 V’s of big data and data science, 2017, [online] Available: https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html. | 15 | In this era of big data, high volumes of a wide variety of valuable data can be easily generated or collected at a high velocity. As such, big data analysis is in demand in various real-life applications and services (e.g., healthcare) as it helps to find useful information and valuable knowledge that are embedded in the big data. The resulting information and knowledge is usually in textual or tabular forms. Given that "a picture is worth a thousand words", visualization and visual analytics helps. In this paper, we present a system for visualizing smart data-as well as their related information and knowledge-from health data, which in turn supports healthcare analytics. | <https://ieeexplore.ieee.org/document/8875352> |
| 25 | M. Banane and M. Ezziyyani, "A Survey on RDF Data Store Based on NoSQL Systems for the Semantic Web Applications" in Advanced Intelligent Systems for Sustainable Development (AI2SD'2018). AI2SD 2018. Advances in Intelligent Systems and Computing, Cham:Springer, vol. 915, 2019. | 26 | The evolution and democratization of technologies have created a veritable explosion of information, this, of course, gives rise to an urgent need to analyze and deal with the huge masses of data. In fact, the problems raised by the accumulation of data (storage, processing time, heterogeneity, capture rate/generation, etc.) become stronger because the data are massive, complex and varied. It is clear that the representation of information has the capacity to synthesize and condense data, and constitutes an efficient approach for analysis. Nevertheless, it remains ineffective and does not solve these problems. Besides that, conventional visualization techniques are rarely adapted to manage and process this mass of information. To face the complex big data challenges, various types of technologies have been developed. This paper talks about the visualization layer. This layer is located just above the Data Sources, Ingestion, Hadoop storage and Hadoop Platform Management layers for which we have already proposed a meta-modeling. It has a very important role at the level of Big Data Structure. In a continuous effort, we shall present in this paper a universal meta-modeling for the visualization layer and its relationship with the other layers of a Big Data system. | <https://ieeexplore.ieee.org/document/8880276> |
| 26 | C. Yao, S. Wu, Z. Liu and P. Li, "A deep learning model for predicting chemical composition of gallstones with big data in medical Internet of Things", *Future Gener. Comput. Syst.*, vol. 94, pp. 140-147, May 2019. | 15 | The digital transformations and use of healthcare information system, electronic medical records, wearable technology, and smart devices are increasing with the passage of time. A variety of sources of big data in healthcare are available, such as biometric data, registration data, electronic health record, medical imaging, patient reported data, biomarker data, clinical data, and administrative data. Visualization of data is a key tool for producing images, diagrams, or animations to convey messages from the viewed insight. The role of cardiology in healthcare is obvious for living and life. The function of heart is the control of blood supply to the entire parts of the body. Recent speedy growth in healthcare and the development of computation in the field of cardiology enable researchers and practitioners to mine and visualize new insights from patient data. The role of visualization is to capture the important information from the data and to visualize it for the easiness of doctors and practitioners. To help the doctors and practitioners, the proposed study presents a detailed report of the existing literature on visualization of data in the field of cardiology. This report will support the doctors and practitioners in decision-making process and to make it easier. This detailed study will eventually summarize the results of the existing literature published related to visualization of data in the cardiology. | <https://ieeexplore.ieee.org/document/8805062> |
| 27 | F. Jeanquartier, C. Jean-Quartier and A. Holzinger, "Visualizing uncertainty for comparing genomic pediatric brain cancer data", *IV 2019 Part I*, pp. 388-391. | 31 | In the current era of big data, a huge amount of data has been generated and collected from a wide variety of rich data sources. Embedded in these big data are useful information and valuable knowledge. An example is healthcare and epidemiological data such as data related to patients who suffered from epidemic diseases like the coronavirus disease 2019 (COVID-19). Knowledge discovered from these epidemiological data helps researchers, epidemiologists and policy makers to get a better understanding of the disease, which may inspire them to come up ways to detect, control and combat the disease. As “a picture is worth a thousand words”, having methods to visualize and visually analyze these big data makes it easily to comprehend the data and the discovered knowledge. In this paper, we present a big data visualization and visual analytics tool for visualizing and analyzing COVID-19 epidemiological data. The tool helps users to get a better understanding of information about the confirmed cases of COVID-19. Although this tool is designed for visualization and visual analytics of epidemiological data, it is applicable to visualization and visual analytics of big data from many other real-life applications and services. | <https://ieeexplore.ieee.org/document/9373130> |
| 28 | R. S. Raghav, Sujatha Pothula, T. Vengattaraman and Dhavachelvan Ponnurangam, "A survey of data visualization tools for analyzing large volume of data in big data platform", *International Conference on Communication and Electronics Systems (ICCES)*, pp. 1-6, 2016. | 0 | The term data analytics is referred to as the study of raw data to gain knowledge. Visualizations are the best way to represent the analyzed data so that it could be consumed by anyone on the go. However, when the amount of data increases and use cases differ, creating these visualizations indulge difficulties and will become repetitive yet hectic tasks. The insights platform depicted here is a generic platform to ease the above said complex process. Here, a generalized platform is made using the internal data from the company and a full stack application is furnished. Since it's developed keeping the generic aspect in mind, the platform could be used for a wide range of use cases by making changes to the data engineering configurations and visualization queries. All others will remain the same and hence new and new applications could be developed in no time with much ease. | <https://ieeexplore.ieee.org/document/9362365> |
| 29 | Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie and S. Gjessing, "Cognitive machine-to-machine communications: Visions and potentials for the smart grid", IEEE Netw., vol. 26, no. 3, pp. 6-13, May 2012. | 53 | Smart grids have been gradually replacing the traditional power grids since the last decade. Such transformation is linked to adding a large number of smart meters and other sources of information extraction units. This provides various opportunities associated with the collected big data. Hence, the triumph of the smart grid energy paradigm depends on the factor of big data analytics. This includes the effective acquisition, transmission, processing, visualization, interpretation, and utilization of big data. The paper provides deep insights into various big data technologies and discusses big data analytics in the context of the smart grid. The paper also presents the challenges and opportunities brought by the advent of machine learning and big data from smart grids. | <https://ieeexplore.ieee.org/document/9272794> |
| 30 | W.-J. Kim, "Knowledge-based diagnosis and prediction using big data and deep learning in precision medicine", *Investigative Clin. Urol.*, vol. 59, pp. 69-71, 2018. | 30 | With the development of information technology, the informationization of the medical industry is also constantly developing rapidly, and medical data is growing exponentially. In the context of “Big Data +”, people began to study the application of data visualization to medical data. Data visualization can make full use of the human sensory vision system to guide users through data analysis and present information hidden behind the data in an intuitive and easy-to-use manner. This paper first introduces the workflow of DBN, a deep learning algorithm, and summarizes the computational characteristics of the algorithm. The classification function is translated into an assembler using an instruction set-based assembly language, and the program is evaluated for performance. Secondly, based on the Hadoop ecosystem, this paper analyzes the BDMISS system for big data medical information resource sharing. Based on the system's requirements and functional positioning, from the medical information collection and sharing, data mining and knowledge management level, the big data medical service system is constructed. Based on the semantic network and ontology theory, big data mining technology and the design of “medical cloud”, the resource sharing mechanism is analyzed. Based on the Spring MVC framework, using Echarts, HCharts and other data visualization technology, according to the design of specific modules, the visualization and display of medical data is realized, which has certain promotion effect on the research and development of medical big data visualization analysis. | <https://ieeexplore.ieee.org/document/8884176> |