

Lapisan Arsitektur Big Data dalam Kajian Studi Pustaka

Ahmad Cucus, Robby Yuli Endra, Yuthsi Aprilinda, Yanuarius Yanu Dharmawan

Program Studi Informatika, Fakultas Ilmu Komputer
Program Studi Bahasa Inggris, Fakultas Ilmu Keguruan dan Pendidikan
Universitas Bandar Lampung
Bandar Lampung, Indonesia

ahmad.cucus@ubl.ac.id, robbi.yulindra@ubl.ac.id, yuthsi.aprilinda@ubl.ac.id, yanu@ubl.ac.id

Abstract- The era of big data is an interesting phenomenon to be discussed by researchers and software developers, application development and data management concepts are increasingly becoming variants and support makes the big data framework can enter every line of life, data that is arranged both synchronously and asynchronously, involves machines and humans in data collection make this technology more in line with the concept of the Industrial Revolution 4.0. In various studies, Big Data concepts and frameworks are presented, from these studies some researchers present layers in the Big Data architecture, where each layer provides input to other layers to be able to be processed into a ready-to-serve form in the community, a layer consisting of data collection, data storage, data processing and data analysis, so that at the application layer the use of data can be felt more optimally by users This paper presents some literature study materials which are summarized to get an explanation of the layers of Big Data architecture that can be developed and applied to other research fields.

Keywords: Big data, Big Data Architecture, Big Data Architecture Layers.

Abstrak- Era big data menjadi sebuah fenomena yang menarik untuk di bahas oleh kalangan peneliti dan pengembang perangkat lunak, pengembangan aplikasi dan konsep pengelolaan data semakin banyak varian dan dukungan menjadikan kerangka big data dapat masuk kesetiap lini kehidupan, data yang tersusun baik secara sinkronus maupun asinkronus, melibatkan mesin dan manusia dalam pengumpulan data menjadikan teknologi ini semakin sejalan dengan konsep Revolusi Industri 4.0 Dalam berbagai kajian di sajikan konsep dan kerangka kerja Big Data, dari kajian tersebut beberapa peneliti menyajikan lapisan dalam arsitektur Big Data, di mana masing masing lapisan memberi input bagi lapisan lain untuk dapat di olah menjadi bentuk yang siap saji di masyarakat, lapisan yang terdiri dari pengumpulan data, penyimpanan data, pemrosesan data serta Analisa data, sehingga pada lapisan aplikasi penggunaan data dapat lebih maksimal di rasakan oleh pengguna. Dalam makalah ini di sajikan beberapa bahan studi literature yang di rangkum untuk mendapatkan penjelasan mengenai lapisan arsitektur Big Data yang dapat di kembangkan dan di terapkan pada bidang penelitian lain.

Kata Kunci: Big data, Arsitektur Big Data, Lapisan Arsitektur Big Data.

1. Pendahuluan

Pada pengembangan ilmu pengetahuan dan teknologi memberikan pengaruh yang sangat besar terhadap kehidupan, makin meningkatnya teknologi penyimpanan data serta teknologi pemrosesan data memberikan keuntungan terhadap berbagai pihak, dalam pengembangan bisnis, pendidikan, kesehatan, politik bahkan dalam kehidupan sosial masyarakat.

Big data merupakan sebuah kajian yang memiliki makna yang cukup besar baik bara bagi para peneliti maupun bagi pengembang teknologi. Karena ranah di dalam pengembangan big data saat ini telah memasuki berbagai lini. Data menjadi bahan mentah berharga bagi dunia bisnis. Dalam bidang ekonomi, ketersediaan data hampir setara dengan modal dan tenaga kerja. Saat ini,

data yang akan dianalisis bersifat dinamis dan memiliki volume yang besar, serta memiliki kelompok data yang berbeda beda. Data saat ini dapat diakses melalui berbagai macam sumber dalam bentuk data asset yang telah banyak tersedia. Oleh karena itu, Big Data memiliki fitur unik seperti heterogen, tidak terstruktur, semi terstruktur, tidak lengkap, berdimensi tinggi [1]

Selain dari berbagai sumber yang dapat kita akses, Data dapat dikumpulkan dalam kecepatan yang tinggi, karena itu penggunaan big data bagi kepentingan dunia pengetahuan dapat dilakukan [2]. Data besar mengacu pada kumpulan data besar yang dicirikan oleh volume yang lebih besar serta variasi dan kompleksitas yang lebih besar, yang dihasilkan pada kecepatan yang lebih tinggi

daripada data operasional normal yang telah ditangani organisasi sebelumnya [3]. Namun, teknologi big data pada saat ini belum sepenuhnya dimanfaatkan karena proses Analisa data telah terhalang oleh tantangan sumber daya manusia, teknologi, hukum, dan etika yang ada. Big data adalah teknologi progresif dan berkembang yang penggunaan yang tepat, serta dapat sebagai alat untuk merevolusi masyarakat secara signifikan [4]. Untuk dapat mengubah Big Data menjadi Nilai nyata yang dapat di gunakan oleh masyarakat, terdapat lima lapisan arsitektur untuk pemrosesan dan analitik data besar (BDPA), termasuk lapisan koleksi, lapisan penyimpanan, lapisan pemrosesan, lapisan analitik, dan lapisan aplikasi [5].

2. Lapisan Arsitektur Big Data

Big data adalah fase berkembang yang berarti volume besar data terstruktur, semi-terstruktur dan tidak terstruktur yang menimbulkan tugas yang sulit untuk diproses menggunakan metode dan database tradisional. Data yang sangat banyak tersebut dapat berasal dari beberapa sumber berbeda seperti sistem transaksi bisnis, database pelanggan, aplikasi seluler, situs web, data yang dihasilkan mesin, dan sensor data waktu nyata yang digunakan dalam lingkungan internet of things (IoT) [6] Juneja 2019 juga mengemukakan ada lima kunci utama kegiatan big data yang dikenal sebagai 5V yaitu Volume, Variety, Velocity, Veracity, Value.

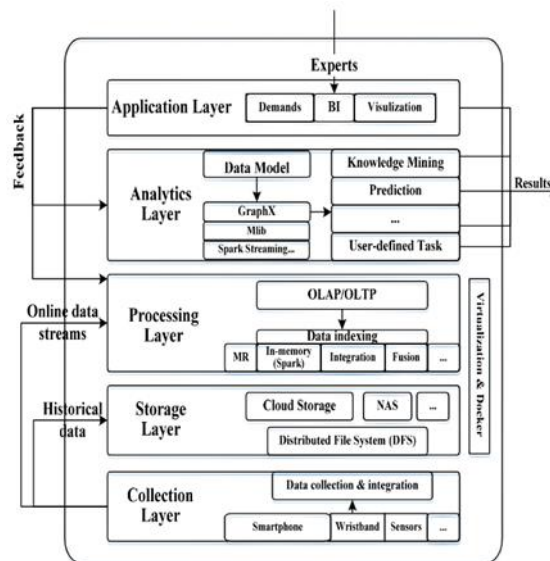
1. Volume: data dalam kumpulan data besar yang dibuat pada tingkat frekuensi tinggi.
2. Variasi: bergagai jenis data yang berbeda, yaitu terstruktur, semi-terstruktur, dan tidak terstruktur
3. Velocity: berkaitan dengan kecepatan dan frekuensi di mana data dapat dihasilkan oleh suatu aplikasi.
4. Kebenaran: berkaitan dengan keakuratan, kebenaran data, dan keasliannya.

5. Nilai: Ini berkaitan dengan kelayakan data yang diekstraksi dari berbagai data mentah yang tersedia.

Dalam proses penggunaan dan pengembangan Big Data, kita juga melihat bagaimana proses dan siklus yang ada dalam big data, secara umum siklus dalam big data dapat di bagi menjadi 5 fase hal ini di kemukakan oleh rashid pada tahun 2019. Adapun fase fase tersebut adalah sebagai berikut [7]

1. Akuisisi Data dan Gudang data, merupakan teknik untuk mengumpulkan data dari berbagai sumber dan menyimpan data.
2. Penambangan dan Pembersihan Data hal ini berkaitan dengan ekstraksi data serta pembersihan data dalam big data
3. Agregasi dan Integrasi, proses ini diperlukan untuk mengumpulkan data serupa dan mengintegrasikan data untuk menghilangkan redundansi dan pengulangan
4. Analisis dan Pemodelan Data, proses dimulai ketika data diperoleh, disimpan, ditambang, dibersihkan dan diintegrasikan.
5. Interpretationis menemukan arti sebenarnya dari pengetahuan yang diekstraksi dan untuk membuat keputusan yang tepat dilakukan dalam proses interpretasi data.

Dari proses serta kunci utama yang telah di sajikan oleh big data, maka big dara memberikan sebuah lapisan arsitektur tersendiri, Zhu pada risetnya di tahun 2019 mengemukakan Arsitektur data besar didorong oleh peningkatan volume data, terutama aliran data besar online, dan oleh tuntutan analisis data untuk semua jenis industri. Arsitektur lapisan untuk BDPA, seperti yang ditunjukkan pada Gambar 1. Lapisan tersebut adalah lapisan kumpulan, lapisan penyimpanan, lapisan pemrosesan, lapisan analitik, dan lapisan aplikasi, yang berlapis dari bawah keatas [5].



Gambar 1. Arsitektur BDPA [5]

3. Lapisan Pengumpulan Data

Lapisan pengumpulan data berfungsi sebagai dasar untuk seluruh solusi BDPA. Dengan semakin banyaknya sumber data, seperti data ponsel cerdas, data jaringan, data sensor, data media sosial, data kesehatan, dll., [5], dalam proses pengumpulan data penggunaan kerumunan seperti pengguna ponsel merupakan hal yang telah banyak dilakukan oleh bisnis pada saat ini [8]

Chen juga mengemukakan ada 2 model dalam proses pengumpulan data terutama data yang di dapat dari komunitas yang melakukan proses pengumpulan data menggunakan perangkat ponsel yaitu secara Asynchronous dan synchronous. Hal ini dilakukan mengingat keterbatasan server dan jaringan data yang mungkin terjadi pada saat pengumpulan data [8]

Sinkronisasi data dapat di implementasi pada sistem terdistribusi dengan kebutuhan ruang data yang besar, selain proses sinkronisasi transfer data, ada dua tahap pekerjaan yang dilakukan dalam sinkronisasi data pada saat data di kumpulkan, Yang pertama adalah memulai sistem saat diinstal. Yang kedua adalah pemulihan sinkronisasi ketika terganggu karena beberapa alasan dan mengganggu terhadap proses pengumpulan data [9]

Tantangan pengumpulan data juga ada pada platform OSN (Online Social Network) dimana platform ini juga menjadi salah satu andalan dalam proses pengumpulan data, Platform OSN, seperti Twitter, Facebook dan LinkedIn, mewakili informasi dalam abstraksi serupa. Setiap pengguna memiliki pengenal unik, Koneksi antara pengguna dapat diambil sebagai daftar tepi, yang menunjukkan koneksi timbal balik (teman) atau langsung (pengikut) antara dua pengguna [10]

Pengumpulan data yang dilakukan selain menggunakan data yang tersedia seperti dalam dataset atau platform OSN, saat ini juga terdapat mekanisme dalam pengumpulan data dengan menggunakan sumber data dari kerumunan orang banyak atau yang disebut dengan crowdsourcing.

Saat ini Crowdsourcing adalah cara yang efektif untuk mengatasi masalah pengumpulan data dengan memanfaatkan banyak orang biasa (CrowdWorker). Alur kerja pada kegiatan crowdsourcing adalah sebagai berikut: pada awalnya seorang pemohon mengajukan masalah, kemudian masalah tersebut diubah menjadi banyak tugas (pertanyaan pertanyaan), dan akhirnya pekerja menyelesaikan tugas yang diberikan kepada mereka dan mereka mendapatkan imbalan [11]

Crowdsourcing digunakan untuk memecahkan berbagai macam masalah, dengan salah satu pendekatan yaitu survey, pola kerja crowdsourcing telah dilakukan oleh berbagai platform dengan mekanisme yang sama, yaitu saasan yang ditetapkan ke pekerja manusia, dan pekerja diberi kompensasi untuk menyelesaikan tugas tersebut. Sejak itu, banyak platform crowdsourcing lainnya telah dikembangkan, dan penelitian tentang crowdsourcing telah berkembang pesat di bidang manajemen data, pembelajaran mesin, dan interaksi komputer manusia. [12]

4. Lapisan Penyimpanan Data

Manajemen data merupakan bagian penting dalam mengembangkan layanan di platform data besar.[13]. Penyimpanan dan pengambilan informasi dalam jumlah besar, serta pencarian kecerdasan dalam banyak data, adalah inti dari konsep Big Data, dan inilah mengapa teknologi ini sangat penting bagi komunitas dan masyarakat TI secara keseluruhan [14]. Lapisan penyimpanan berfungsi sebagai lapisan kedua dan menyediakan infrastruktur penyimpanan dasar yang dapat diperluas untuk penyimpanan data historis. Arsitektur BDPA memerlukan teknologi penyimpanan yang memiliki karakteristik berikut: dapat diskalakan, dengan penyimpanan berjenjang, pengelolaan mandiri, sangat tersedia, dapat diakses secara luas, dan dapat dipulihkan [5]

Media penyimpanan saat ini mengalami perkembangan teknologi yang cukup pesat baik dalam skala maupun kualitas, Sistem penyimpanan modern menggunakan HDD dan SSD terutama untuk masing-masing mencapai kapasitas besar dan kinerja tinggi [15]. Selain secara fisik tersimpan konsep penggunaan media penyimpanan berbasis cloud juga telah dikembangkan untuk konsep penyimpanan big data [14]

Penyedia media penyimpanan saat ini juga memberikan berbagai kemudahan dalam proses pengolahan data, penggunaan SQL yang lebih complex [16]. serta penggunaan no SQL dengan bermacam platform yang disediakan [17]

5. Lapisan Pemrosesan Data

Lapisan pemrosesan data adalah inti dalam sistem BDPA. Ini adalah lapisan fundamental dari banyak tugas analitik data. Di era big data, sumber data memiliki struktur dan konten yang sangat heterogen [5]. Ada banyak teknik dan banyak cara untuk memproses data besar yang dibawa dari berbagai sumber dalam berbagai format [18]

Dalam layer pemrosesan data secara garis besar di bagi menjadi tahap processing dan pre-processing, Pemrosesan awal data adalah pendekatan penambangan data yang digunakan untuk mengubah data mentah dalam format yang efisien dan paling berguna [18]

Dalam kegiatan tahapan pre pemrosesan terdapat beberapa tahapan yang dilakukan, yaitu data cleaning, data integration, data filtering dan data transformation [5][18]

Pada tahap Data Cleaning proses yang dilakukan mendeteksi dan mengoreksi (atau menghapus) catatan yang rusak atau tidak akurat dari kumpulan catatan, tabel, atau database [19] tahapan data cleaning dilakukan dengan berbagai macam cara seperti verifikasi dan Validasi Data, Proses ini sangat penting dalam proses pengembangan perangkat lunak [20]. Data validasi banyak digunakan untuk mengukur performa generalisasi model dalam pembelajaran mesin, yang selanjutnya akan di proses melalui integrasi data [21] validasi data adalah proses mengidentifikasi data yang mengandung kesalahan (misalnya, nilai di luar jangkauan, fitur tidak ada, dan sebagainya) [22]

Setelah tahap cleaning data, tahap integrasi data juga menjadi tujuan utama dari layer pemrosesan data, mengintegrasikan data yang dikumpulkan dari sumber data terstruktur dan tidak terstruktur [23] setelah proses integrasi selesai maka data memiliki banyak noise yang terdiri dari garbage data, irrelevant data hingga advertisement data, kita dapat melakukan filter data untuk mendapatkan data data yang kita butuhkan. [24].

6. Lapisan Analisis Data

Sistem BDPA memerlukan campuran alat analitik data untuk kebutuhan pengguna yang berbeda [5] karena pada hari ini dunia Industri mulai fokus dan mencoba untuk mengetahui tentang penggunaan Big Data secara efisien, untuk kegiatan analisis dan mengidentifikasi wawasan bisnis yang bermanfaat untuk keputusan bisnis yang lebih baik dan layak [25].

Big Data mengumpulkan data dari Semua sumber yang berbeda menciptakan jenis informasi yang berbeda yang dapat secara komprehensif mendelegasikan informasi Terstruktur, Tidak Terorganisir dan Semi-terorganisir [26] sehingga menjadikan volume besar data terstruktur, semi-terstruktur, dan tidak terstruktur yang dapat digunakan untuk analisis data menggunakan berbagai metode [25]. Bisnis dan Organisasi menggunakan sejumlah besar data tidak terstruktur, terstruktur, atau semi-terstruktur dalam mengelola operasi sehari-hari dan pengambilan keputusan. Oleh karena itu, untuk membuat keputusan yang tepat, diperlukan mekanisme analitik data besar yang tepat [4].

Mekanisme dalam pengolahan dan Analisa data dilakukan dengan berbagai metode, framework maupun Algoritma. Penggunaan mesin learning serta Algoritma Natural Language Processing [27] [28] atau Neural Network [29] dapat digunakan untuk data yang terstruktur maupun semi – terstruktur [30]

Gudang data biasanya menyimpan data bisnis terstruktur. Namun, sebagian besar data dalam organisasi adalah data tidak terstruktur, yang menempati sekitar 80 persen dari data perusahaan [3]. Pengelolaan data data tidak terstruktur juga menjadi salah satu konsen bagi para peneliti, data tidak terstruktur tidak hanya terdiri dari text saja, terdapat pula dalam bentuk image, video, suara dan lainnya. Penggunaan Deep Learning telah terbukti berhasil dalam banyak tugas computer vision mulai dari pengenalan dan deteksi objek hingga segmentasi semantic. [31].

7. Lapisan Aplikasi (Implementasi Big Data)

Lapisan aplikasi bertindak sebagai lapisan tertinggi. Dengan empat lapisan sebelumnya, sistem BDPA dapat membangun berbagai aplikasi untuk pengguna yang berbeda [5] Penggunaan big data dalam berbagai bidang telah banyak di terapkan.

Penggunaan dalam bidang Kesehatan [32] [33] maupun dalam rangka menyikapi pandemi Covid yang terjadi saat ini [2] Pada bidang bisnis [34] dan bidang pendidikan [35] [36] Keluaran dari aplikasi ini membantu pengguna untuk membuat keputusan yang cepat [5].

8. Kesimpulan

Sejak di kenalnya penyimpanan data dengan media digital perkembangan inovasi dan teknologi di bidang pengolahan data berkembang dengan pesat, saat ini kita telah memasuki Era big data di mana teknologi penyimpanan konvensional tidak lagi mampu menangani data yang memiliki volume, velocity dan variety yang sangat besar [37] Perkembangan data seiring dengan perkembangan teknologi informasi yang semakin cepat, data yang besar diproduksi setiap menit, data data tersebut berasal dari berbagai macam industry seperti internet, jaringan komunikasi, Kesehatan dan banyak lainnya [38]

Lapisan arsitektur big data yang terdiri dari kegiatan pengumpulan data, penyimpanan data, pemrosesan data, dan Analisa data akan menentukan keluaran dalam lapisan aplikasi, setiap lapisan memiliki kerangka kerja dan ranah penelitian yang luas dan dalam. Lapisan arsitektur dari big data mempermudah kita dalam menentukan metode dan pendekatan yang tepat untuk menghasilkan luaran yang diinginkan dalam lapisan aplikasi.

9. Daftar Pustaka

- [1] IEEE Staff, *2017 International Conference on Information Communication and Embedded Systems (ICICES)*. IEEE, 2017.
- [2] C. K. Leung, Y. Chen, S. Shang, and D. Deng, "Big Data Science on COVID-19 Data," in *Proceedings - 2020 IEEE 14th International Conference on Big Data Science and Engineering, BigDataSE 2020*, Dec. 2020, pp. 14–21, doi: 10.1109/BigDataSE50710.2020.00010.
- [3] O. Baker and C. N. Thien, "A New Approach to Use Big Data Tools to Substitute Unstructured Data Warehouse," in *2020 IEEE Conference on Big Data and Analytics, ICBDA 2020*, Nov. 2020, pp. 26–31, doi: 10.1109/ICBDA50157.2020.9289757.
- [4] N. Aljehane, "Big Data Analytics: Challenges and Opportunities," Sep. 2020, doi: 10.1109/ICCIT-144147971.2020.9213765.
- [5] J. Y. Zhu, B. Tang, and V. O. K. Li, "A five-layer architecture for big data processing and analytics A five-layer architecture for big data processing and analytics 39," 2019. [Online]. Available: <https://github.com/amplab-extras/>.
- [6] A. Juneja and N. N. Das, "Big Data Quality Framework: Pre-Processing Data in Weather Monitoring Application," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Perspectives Prospect. Com.* 2019, pp. 559–563, 2019, doi: 10.1109/COMITCon.2019.8862267.
- [7] K. L. M. Ang, F. L. Ge, and K. P. Seng, "Big Educational Data Analytics: Survey, Architecture and Challenges," *IEEE Access*, vol. 8. Institute of Electrical and Electronics Engineers Inc., pp. 116392–116414, 2020, doi: 10.1109/ACCESS.2020.2994561.
- [8] W. Chen, Z. Li, Y. Liang, J. Chen, and W. Zhu, "An Asynchronous Distributed Data Collection Approach for Mobile Group Consumption," in

- Proceedings - 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things, IIKI 2015*, Mar. 2016, pp. 19–24, doi: 10.1109/IIKI.2015.11.
- [9] A. S. Luchinin, I. V. Malygin, S. I. Starikov, and M. V. Markov, “Synchronization of elements of the system for collecting data separated by long distances,” *2019 Syst. Signal Synchronization, Gener. Process. Telecommun. SYNCHROINFO 2019*, pp. 1–4, 2019, doi: 10.1109/SYNCHROINFO.2019.8813921.
- [10] IEEE Computer Society, IEEE Computer Society. Technical Committee on the Internet, and Institute of Electrical and Electronics Engineers., *2016 IEEE 2nd International Conference on Collaboration and Internet Computing: IEEE CIC 2016 : proceedings : 1-3 November 2016, Pittsburgh, Pennsylvania, USA.*
- [11] C. Shan, N. Mamoulis, G. Li, R. Cheng, Z. Huang, and Y. Zheng, “A Crowdsourcing Framework for Collecting Tabular Data,” 2018.
- [12] Y. Roh, G. Heo, and S. E. Whang, “A Survey on Data Collection for Machine Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. PP, no. c, p. 1, 2019, doi: 10.1109/TKDE.2019.2946162.
- [13] M. C. Nguyen and H. S. Won, “Data storage adapter in big data platform,” in *Proceedings - 8th International Conference on Database Theory and Application, DTA 2015*, Mar. 2016, pp. 6–9, doi: 10.1109/DTA.2015.9.
- [14] A. Erraissi, M. Banane, A. Belangour, and M. Azzouazi, “Big Data Storage using Model Driven Engineering: From Big Data Meta-model to Cloudera PSM meta-model,” Oct. 2020, doi: 10.1109/ICDABI51230.2020.9325674.
- [15] N. Kenji, K. Joichiro, and Y. Saneyasu, “A Study on Big Data I/O Performance with Modern Storage Systems,” in *017 IEEE International Conference on Big Data (BIGDATA)*, 2017, pp. 278–289, doi: 10.1109/HPCA.2011.5749736.
- [16] K. Evdokia, Konstantinou, and K. Nectarios, “Towards a Multi-engine Query Optimizer for Complex SQL Queries on Big Data,” *2019 IEEE Int. Conf. Big Data (Big Data)*, vol. 59, no. 11, pp. 56–65, Nov. 2019, doi: 10.1145/2934664.
- [17] Y. Man, L. Kui, Z. Liwei, and Z. Chenhong, “Research on Big Data Storage Model of Oilfield Assay Data Based on MongoDB,” 2018.
- [18] V. Desai and H. A. Dinesha, “A Hybrid Approach to Data Pre-processing Methods,” Nov. 2020, doi: 10.1109/INOCON50539.2020.9298378.
- [19] C. Xie, J. Gao, and C. Tao, “Big data validation case study,” in *Proceedings - 3rd IEEE International Conference on Big Data Computing Service and Applications, BigDataService 2017*, Jun. 2017, pp. 281–286, doi: 10.1109/BigDataService.2017.44.
- [20] K. Thiruthanigesan and N. Thiruchchelvan, “Data Verification and Validation Process in the Management System Development,” *Middle-East J. Sci. Res.*, vol. 25, no. 5, pp. 902–911, 2017, doi: 10.5829/idosi.mejsr.2017.902.911.
- [21] T. Takase, “Evaluation of Stratified Validation in Neural Network Training with Imbalanced Data,” 2019.
- [22] Y. Chung, T. Kraska, N. Polyzotis, K. H. Tae, and S. E. Whang, “Automated Data Slicing for Model Validation: A Big Data - AI Integration Approach,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 12, pp. 2284–2296, Dec. 2020, doi: 10.1109/TKDE.2019.2916074.
- [23] M. E. Vidal and S. Jozashoori, “Semantic data integration techniques for transforming big biomedical data into actionable knowledge,” in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, Jun. 2019, vol. 2019-June, pp. 563–566, doi: 10.1109/CBMS.2019.00116.
- [24] Il-Kyu Ha and Bong-Hyun Bac, “Effective Garbage Data Filtering Algorithm for SNS Big Data Processing by Machine Learning,” 2020.
- [25] Suyash Mishra, “Structured and Unstructured Big Data Analytics,” 2017.
- [26] S. Sakineti, “Protagonist of Big Data and Predictive Analytics using data analytics,” 2018.
- [27] R. Devarakonda, M. Giansiracusa, and J. Kumar, “Machine Learning and Social Media to Mine and Disseminate Big Scientific Data,” in *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, Jan. 2019, pp. 5312–5315, doi: 10.1109/BigData.2018.8622470.
- [28] M. Elsayed, A. Abdelwahab, and H. Ahdelkader, “A Proposed Framework for Improving Analysis of Big Unstructured Data in Social Media,” 2019.
- [29] C. Costa, A. Charalampous, A. Konstantinidis, D. Zeinalipour-Yazti, and M. F. Mokbel, “TBD-DP: Telco big data visual analytics with data postdiction,” in *Proceedings - IEEE International Conference on Mobile Data Management*, Jul. 2018, vol. 2018-June, pp. 280–281, doi: 10.1109/MDM.2018.00050.
- [30] K. Ghane, “Big data pipeline with ML-based and crowd sourced dynamically created and maintained columnar data warehouse for structured and unstructured big data,” in *Proceedings - 3rd International Conference on Information and Computer Technologies, ICICT 2020*, Mar. 2020, pp. 60–67, doi: 10.1109/ICICT50521.2020.00018.
- [31] S. Roy *et al.*, “Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound,” *IEEE Trans. Med. Imaging*, vol. 39, no. 8, pp. 2676–2687, Aug. 2020, doi: 10.1109/TMI.2020.2994459.
- [32] M. Panda, “Big Data in Health Care: A Mobile Based Solution,” 2017.
- [33] W. Xing and Y. Bei, “Medical Health Big Data Classification Based on KNN Classification Algorithm,” *IEEE Access*, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.
- [34] S. Savanur and K. S. Shreedhara, “Automated data validation for data warehouse testing,” *2016 Int. Conf. Electr. Electron. Commun. Comput. Optim. Tech. ICEECCOT 2016*, pp. 223–226, 2017, doi: 10.1109/ICEECCOT.2016.7955219.
- [35] N. Alexandrov and V. Alexandrov, “Computational science research methods for science education at

- PG level,” *Procedia Comput. Sci.*, vol. 51, no. 1, pp. 1685–1693, 2015, doi: 10.1016/j.procs.2015.05.305.
- [36] Y. Tan, “How to Effectively Infiltrate Emotional Education in Primary School Chinese Teaching from Perspective of Big Data,” in *Proceedings - 2020 International Conference on Computers, Information Processing and Advanced Education, CIPAE 2020*, Oct. 2020, pp. 138–141, doi: 10.1109/CIPAE51077.2020.00044.
- [37] F. Z. Benjelloun, A. A. Lahcen, and S. Belfkih, “An overview of big data opportunities, applications and tools,” May 2015, doi: 10.1109/ISACV.2015.7105553.
- [38] J. Li, Z. Xu, Y. Jiang, and R. Zhang, “The overview of big data storage and management,” in *Proceedings of 2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2014*, Oct. 2014, pp. 510–513, doi: 10.1109/ICCI-CC.2014.6921508.