Time Series Database (TSD): A Approach for Time Series Data Management

RAJU SHANMUGAM[1],MARJAN KUCHAKI RAFSANJANI [2]
[1]Karnavati University, Gandhinagar, Gujarat, India (e-mail: dr.sraju@sraju.in)
[2]Shahid Bahonar University of Kerman, Kerman, Iran, (e-mail: kuchaki@uk.ac.ir)

ABSTRACT To keep tabs on things throughout time, there is a special kind of database called a time series database (TSD). In a nutshell, time series data are any observations or occurrences that have been observed, sampled, downsampled, and aggregated over time. Many other kinds of analytics data fall under this category, including but not limited to data generated from smart devices. To better manage time-stamped metrics, events, and measurements, a database designed for time series is necessary. For tracking shifts over time, a TSD is the most accurate tool available. In this context, we analyze the development in the field of TSD. We used the Scopus database to collect the relevant papers.

KEYWORDS Time Series; IoT; Scopus

I. INTRODUCTION

Although not a novel concept, time series databases have historically been used to analyze financial data, market volatility, and many more problems [1]–[3]. The financial sector, although still a major user of time series data, is only one of many examples of where this kind of information is put to use today. Over the last decade, there have been significant shifts in the environment under which computers operate; due to development in the field of IoT and other smart devices [4]–[7]. Things have gotten really specialized. The age of the component has arrived, and with it, the age of the component. We are also seeing the instrumentation of every conceivable surface in the material universe. A sensor may be found on or is being built into almost every object [8]–[11]. Consequently, the business now generates a continuous flow of measurements, events, and time series data from all sources, both within and outside the organization. This necessitates a shift in the underlying platforms to accommodate the increased data density, data variety, monitoring requirements, and control requirements of these new workloads [12]–[14]. The current era necessitates, and indeed is causing, a radical change in the way we conceptualize and implement data infrastructure, as well as how we construct, monitor, command, and administer complex systems [15]–[18]. We need a time series database that is fast, scalable, and designed specifically for time series data [19], [20].

II. RESEARCH METHODOLOGY

In this article, we analyze the development in the field of time series databases. We search the Scopus database using the following query:

TITLE-ABS-KEY (timeseries AND database)

III. RESULTS AND DISSCUSSION

In this research, we analyze the work in the field of time series databases. As explained in the previous section, we used the Scopus database to conduct our research. After running the query, we get 117 documents as represented in Figure 1. Figure 2 presents the number of papers published over the time-span related to time series databases. From Figure 1 it is clear that the average growth rate of papers over that year is 10.91%, this shows that TSD is a relevant topic and needs further research. The collected documents are published at different platforms as represented in Figure 2. however, it is clear that the majority of the articles are published in international journals (47.9In addition to that, from Figure 3 it is clear that the majority of computer science researchers are working in the field of exploratory data analysis.

A. AUTHOR ANALYSIS

Within this subsection, we conduct an examination of the composition of the authors. We sort the writers by the total number of articles they have contributed to. This arrangement is represented as follows:

- ANDERSON (102)
- ATKINSON (102)
- KANG S (102)
- MILLS IC (102)
- WALTON H (102)
- ANDERSSON M (85)
- EKLUND A (85)
- JOHANNESSON M (85)

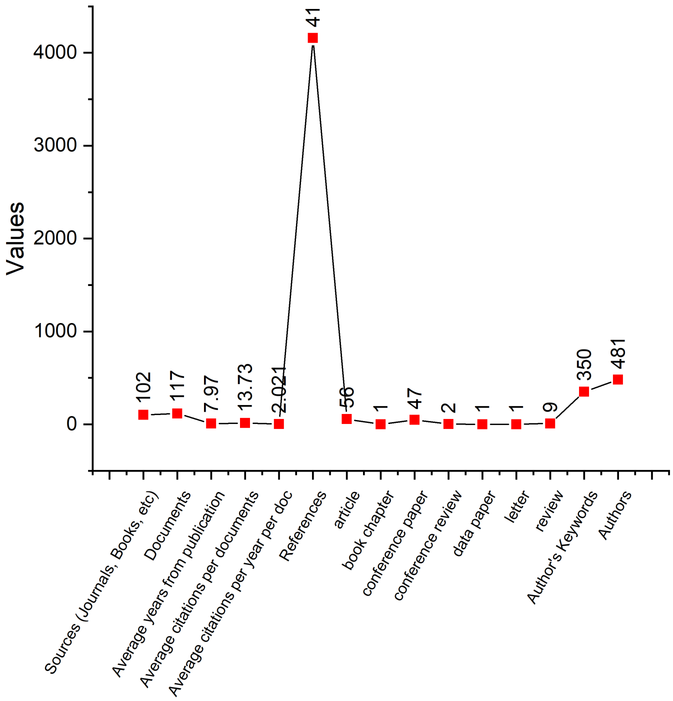

FIGURE 1: General Information

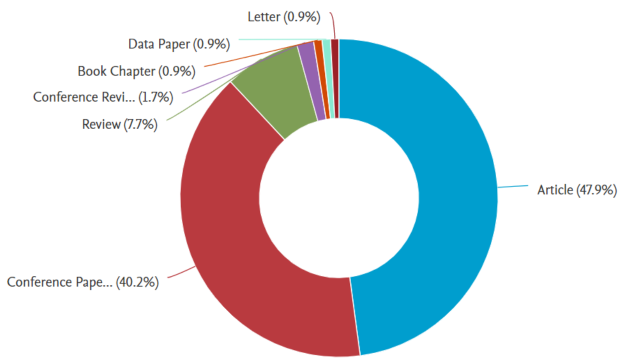

FIGURE 2: Different Type

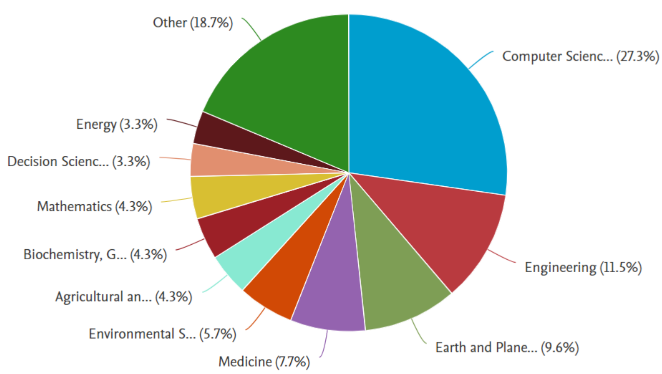

FIGURE 3: Different Domain

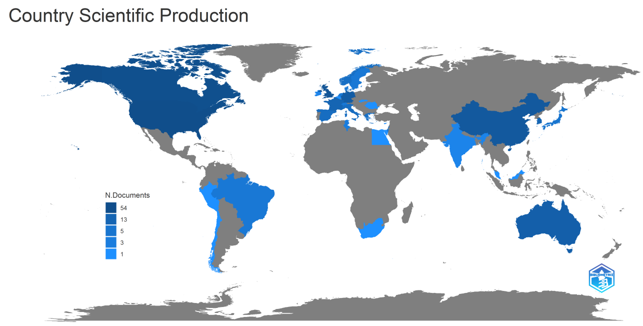

FIGURE 4: Different Country Distribution

- JOSEPHSON C (85)
- KNUTSSON H (85)

This subsection helps us to find the leading researcher in the field of TSD. This will help young researchers to find relevant articles on TSD.

B. ANALYSIS OF COUNTRY PRODUCTION

Researchers' locations matter much to the development of scientific inquiry, too. As such, this section assesses how the location influences TSD studies. We show our findings in Figure 4. The quantity of articles produced by a country's researchers may be used to rank the nations. From Figure 4; the ranking of the countries are as follows.

- USA (54)
- CANADA (49)
- UK (35)
- CHINA (27)
- GERMANY (22)
- AUSTRALIA (18)
- FRANCE (13)
- SOUTH KOREA (13)

C. ANALYSIS OF DOCUMENT DISTRIBUTION

This section provides an overview of how the articles were disseminated across academics. We analyzed data from 117 sources that were indexed in Scopus. These sources included journals, conference papers, reviews, books, and book chapters. The most-cited works in TSD give a summary of the field's most pressing questions and most consequential concepts.Table 1 presents the distribution of the paper according to the total number of citations

D. ANALYSIS OF KEYWORD DISTRIBUTION

We all know that keywords may be used to provide a summary of the studies that were conducted. Consequently, a concise picture of the research area may be gleaned from an analysis of the Scopus database's keyword distribution.Figure 5 presents the keyword distribution. In the Figure 5. From Figure 5 it is clear that frequently occurring keywords are as follows:

- timeseries (6)

TABLE 1: Highly Cited Papers

Paper	DOI	Total Citations
MILLS IC, 2015, BMJ OPEN [21]	10.1136/bmjopen-2014-006946	102
EKLUND A, 2012, NEUROIMAGE [22]	10.1016/j.neuroimage.2012.03.093	85
GALBRAITH ED, 2013, NAT GEOSCI [23]	10.1038/ngeo1832	75
KLEE U, 2006, EURASIP J APPL SIGN PROCESS [24]	10.1155/ASP/2006/12378	70
DAJANI DR, 2016, AUTISM RES [25]	10.1002/aur.1494	68
ANDERSEN MP, 2016, PROC USENIX CONF FILE STORAGE TECHNOL , FAST [26]	NA	55
ALEXANDER LV, 2020, ENVIRON RES LETT [27]	10.1088/1748-9326/ab79e2	53
HOLT J, 2012, PROG OCEANOGR [28]	10.1016/j.pocean.2012.08.001	48
LIN J, 2012, ADVANCES IN MACHINE LEARNING AND DATA MIN FOR ASTRONOMY [29]	10.1201/b11822-22	47
CHATTERJEE A, 2020, SENSORS [30]	10.3390/s20113089	46
PEINGS Y, 2010, CLIM DYN [31]	10.1007/s00382-009-0565-0	46
LAWES T, 2012, BMJ OPEN [32]	10.1136/bmjopen-2011-000797	46
TERENZIANI P, 2003, IEEE TRANS KNOWL DATA ENG [33]	10.1109/TKDE.2003.1185847	35
CASH BA, 2017, CLIM DYN [34]	10.1007/s00382-016-3320-3	34
WAX TM, 1977, JOURNAL OF COMPARATIVE AND PHYSIOLOGICAL PSYCHOLOGY [35]	10.1037/h0078071	33
BLAAS J, 2009, IEEE TRANS VISUAL COMPUT GRAPHICS [36]	10.1109/TVCG.2009.181	32
XU X, 2014, PROC - IEEE INT CONF WEB SERV , ICWS [37]	10.1109/ICWS.2014.45	29
MONTI C, 2013, PROC INT WORKSHOP ISSUES SENTIMENT DISCOV OPIN MIN , WISDOM - HELD CONJUNCTION SIGKDD [38]	10.1145/2502069.2502072	28
FLIS A, 2015, OPEN BIOL [39]	10.1098/rsob.150042	27
LEE AJT, 2009, DATA KNOWL ENG [40]	10.1016/j.datak.2009.04.005	27

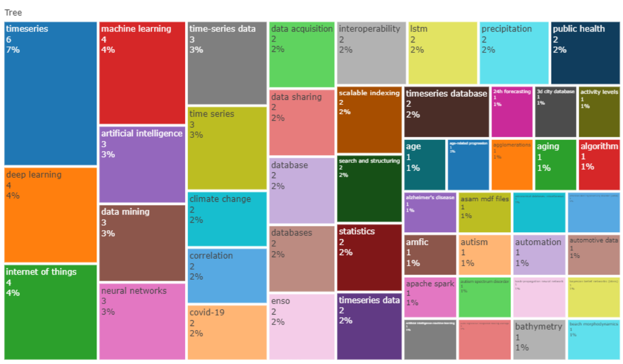

FIGURE 5: Keyword Distribution

- deep learning (4)
- internet of things (4)
- machine learning (4)
- artificial intelligence (3)
- data mining (3)
- neural networks (3)
- time-series data (3)

IV. CONCLUSION

Data records that make up a 'time series,' or a group of data points with timestamps, may be stored and retrieved with the use of a computer system called a time-series database (TSD). Each data point's relationship to the others is greatly enhanced by the inclusion of a timestamp. Time-series data, including sensor readings and intraday stock prices, often comes in the form of a constant stream. For sophisticated analysis of time-stamped data, a time-series database enables for quick data entry and retrieval. In this context, this paper helps the research to get a better understanding of TSD.

REFERENCES

[1] B. B. Gupta, S. Gupta, and P. Chaudhary, "Enhancing the browser-side context-aware sanitization of suspicious html5 code for halting the dom-based xss vulnerabilities in cloud," International Journal of Cloud Applications and Computing (IJCAC), vol. 7, no. 1, pp. 1–31, 2017.

[2] N. Kumar and et al., "A novel framework for risk assessment and resilience of critical infrastructure towards climate change," Technological Forecasting and Social Change, vol. 165, p. 120532, 2021.

[3] A. Gaurav, V. Arya, and D. Santaniello, "Analysis of machine learning based ddos attack detection techniques in software defined network," Cyber Security Insights Magazine (CSIM), vol. 1, no. 1, pp. 1–6, 2022.

[4] A. Tewari and et al., "A lightweight mutual authentication protocol based on elliptic curve cryptography for iot devices," International Journal of Advanced Intelligence Paradigms, vol. 9, no. 2-3, pp. 111–121, 2017.

[5] A. Tewari and et al., "Secure timestamp-based mutual authentication protocol for iot devices using rfid tags," International Journal on Semantic Web and Information Systems (IJSWIS), vol. 16, no. 3, pp. 20–34, 2020.

[6] M. Kaur and et al., "Secure and energy efficient-based e-health care framework for green internet of things," IEEE Transactions on Green Communications and Networking, vol. 5, no. 3, pp. 1223–1231, 2021.

[7] B. B. Gupta, S. Yamaguchi, and D. P. Agrawal, "Advances in security and privacy of multimedia big data in mobile and cloud computing," Multimedia Tools and Applications, vol. 77, no. 7, pp. 9203–9208, 2018.

[8] L. Zou and et al., "A novel coverless information hiding method based on the average pixel value of the sub-images," Multimedia tools and applications, vol. 78, no. 7, pp. 7965–7980, 2019.

[9] C. L. Stergiou and et al., "Secure machine learning scenario from big data in cloud computing via internet of things network," in Handbook of computer networks and cyber security. Springer, 2020, pp. 525–554.

[10] A. Singh and et al., "Distributed denial-of-service (ddos) attacks and defense mechanisms in various web-enabled computing platforms: Issues, challenges, and future research directions," International Journal on Semantic Web and Information Systems (IJSWIS), vol. 18, no. 1, pp. 1–43, 2022.

[11] P. Do and et al., "Building a knowledge graph by using cross-lingual transfer method and distributed minie algorithm on apache spark," Neural Computing and Applications, pp. 1–17, 2020.

[12] M. Hammad and et al., “Myocardial infarction detection based on deep neural network on imbalanced data,” Multimedia Systems, vol. 28, no. 4, pp. 1373–1385, 2022.

[13] A. Dahiya and et al., “A reputation score policy and bayesian game theory based incentivized mechanism for ddos attacks mitigation and cyber defense,” Future Generation Computer Systems, vol. 117, pp. 193–204, 2021.

[14] M. Casillo and et al., “Context aware recommender systems: A novel approach based on matrix factorization and contextual bias,” Electronics, vol. 11, no. 7, p. 1003, 2022.

[15] B. Gupta, P. Chaudhary, X. Chang, and N. Nedjah, “Smart defense against distributed denial of service attack in iot networks using supervised learning classifiers,” Computers & Electrical Engineering, vol. 98, p. 107726, 2022.

[16] R. Jiao and et al., “Adaptive feature selection and construction for day-ahead load forecasting use deep learning method,” IEEE Transactions on Network and Service Management, vol. 18, no. 4, pp. 4019–4029, 2021.

[17] B. B. Gupta, A. Gaurav, P. K. Panigrahi, and V. Arya, “Analysis of artificial intelligence-based technologies and approaches on sustainable entrepreneurship,” Technological Forecasting and Social Change, vol. 186, p. 122152, 2023.

[18] S. Kumar, S. Kumar, N. Ranjan, S. Tiwari, T. R. Kumar, D. Goyal, G. Sharma, V. Arya, and M. K. Rafsanjani, “Digital watermarking-based cryptosystem for cloud resource provisioning,” International Journal of Cloud Applications and Computing (IJCAC), vol. 12, no. 1, pp. 1–20, 2022.

[19] R. K. S. Rajput, D. Goyal, A. Pant, G. Sharma, V. Arya, and M. K. Rafsanjani, “Cloud data centre energy utilization estimation: Simulation and modelling with idr,” International Journal of Cloud Applications and Computing (IJCAC), vol. 12, no. 1, pp. 1–16, 2022.

[20] K. Pathoee, D. Rawat, A. Mishra, V. Arya, M. K. Rafsanjani, and A. K. Gupta, “A cloud-based predictive model for the detection of breast cancer,” International Journal of Cloud Applications and Computing (IJCAC), vol. 12, no. 1, pp. 1–12, 2022.

[21] I. Mills, R. Atkinson, S. Kang, H. Walton, and H. Anderson, “Quantitative systematic review of the associations between short-term exposure to nitrogen dioxide and mortality and hospital admissions,” BMJ Open, vol. 5, no. 5, 2015.

[22] A. Eklund, M. Andersson, C. Josephson, M. Johannesson, and H. Knutsson, “Does parametric fmri analysis with spm yield valid results?- an empirical study of 1484 rest datasets,” NeuroImage, vol. 61, no. 3, pp. 565–578, 2012.

[23] E. Galbraith, M. Kienast, A. Albuquerque, M. Altabet, F. Batista, D. Bianchi, S. Calvert, S. Contreras, X. Crosta, R. De Pol-Holz, N. Dubois, J. Etourneau, R. Francois, T.-C. Hsu, T. Ivanochko, S. Jaccard, S.-J. Kao, T. Kiefer, S. Kienast, M. Lehmann, P. Martinez, M. McCarthy, A. Meckler, A. Mix, J. Möbius, T. Pedersen, L. Pichevin, T. Quan, R. Robinson, E. Ryabenko, A. Schmittner, R. Schneider, A. Schneider-Mor, M. Shigemitsu, D. Sinclair, C. Somes, A. Studer, J.-E. Tesdal, R. Thunell, and J.-Y. Terence Yang, “The acceleration of oceanic denitrification during deglacial warming,” Nature Geoscience, vol. 6, no. 7, pp. 579–584, 2013.

[24] U. Klee, T. Gehrig, and J. McDonough, “Kalman filters for time delay of arrival-based source localization,” Eurasip Journal on Applied Signal Processing, vol. 2006, pp. 1–15, 2006.

[25] D. Dajani and L. Uddin, “Local brain connectivity across development in autism spectrum disorder: A cross-sectional investigation,” Autism Research, vol. 9, no. 1, pp. 43–54, 2016.

[26] M. Andersen and D. Culler, “Btrdb: Optimizing storage system design for timeseries processing,” 2016, pp. 39–52.

[27] L. Alexander, M. Bador, R. Roca, S. Contractor, M. Donat, and P. Nguyen, “Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products,” Environmental Research Letters, vol. 15, no. 5, 2020.

[28] J. Holt, S. Hughes, J. Hopkins, S. Wakelin, N. Penny Holliday, S. Dye, C. González-Pola, S. Hjøllo, K. Mork, G. Nolan, R. Proctor, J. Read, T. Shammon, T. Sherwin, T. Smyth, G. Tattersall, B. Ward, and K. Wiltshire, “Multi-decadal variability and trends in the temperature of the northwest european continental shelf: A model-data synthesis,” Progress in Oceanography, vol. 106, pp. 96–117, 2012.

[29] J. Lin, S. Williamson, K. Borne, and D. DeBarr, Pattern Recognition in Time Series, 2012.

[30] A. Chatterjee, M. Gerdes, and S. Martinez, “Statistical explorations and univariate timeseries analysis on covid-19 datasets to understand the trend of disease spreading and death,” Sensors (Switzerland), vol. 20, no. 11, 2020.

[31] Y. Peings and H. Douville, “Influence of the eurasian snow cover on the indian summer monsoon variability in observed climatologies and cmip3 simulations,” Climate Dynamics, vol. 34, no. 5, pp. 643–660, 2010.

[32] T. Lawes, B. Edwards, J.-M. López-Lozano, and I. Gould, “Trends in staphylococcus aureus bacteraemia and impacts of infection control practices including universal mrsa admission screening in a hospital in scotland, 2006-2010: Retrospective cohort study and time-series intervention analysis,” BMJ Open, vol. 2, no. 3, 2012.

[33] P. Terenziani, “Symbolic user-defined periodicity in temporal relational databases,” IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 2, pp. 489–509, 2003.

[34] B. Cash, R. Barimalala, I. Kinter, J.L., E. Altshuler, M. Fennessy, J. Manganello, F. Molteni, P. Towers, and F. Vitart, “Sampling variability and the changing enso–monsoon relationship,” Climate Dynamics, vol. 48, no. 11-12, pp. 4071–4079, 2017.

[35] T. Wax and W. Teena M., “Effects of age, strain, and illumination intensity on activity and self-selection of light-dark schedules in mice,” Journal of Comparative and Physiological Psychology, vol. 91, no. 1, pp. 51–62, 1977.

[36] J. Blaas, C. Botha, E. Grundy, M. Jones, R. Laramee, and F. Post, “Smooth graphs for visual exploration of higher-order state transitions,” IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 969–976, 2009.

[37] X. Xu, S. Huang, Y. Chen, K. Brown, I. Halilovic, and W. Lu, “Tsaaas: Time series analytics as a service on iot,” 2014, pp. 249–256.

[38] C. Monti, A. Rozza, G. Zappella, M. Zignani, A. Arvidsson, and E. Colleoni, “Modelling political disaffection from twitter data,” 2013.

[39] A. Flis, A. Fernández, T. Zielinski, V. Mengin, R. Sulpice, K. Stratford, A. Hume, A. Pokhilko, M. Southern, D. Seaton, H. McWatters, M. Stitt, K. Halliday, and A. Millar, “Defining the robust behaviour of the plant clock gene circuit with absolute rna timeseries and open infrastructure,” Open Biology, vol. 5, no. 10, 2015.

[40] A. Lee, H.-W. Wu, T.-Y. Lee, Y.-H. Liu, and K.-T. Chen, “Mining closed patterns in multi-sequence time-series databases,” Data and Knowledge Engineering, vol. 68, no. 10, pp. 1071–1090, 2009.