Temporal Aspects of Big Data Management: State-of-the-Art Analysis and Future Research Directions

Alfredo Cuzzocrea
University of Trieste and ICAR-CNR, Italy
alfredo.cuzzocrea@dia.units.it

***Abstract*—A great deal of research efforts has been invested in *temporal aspects of big data management* during last years, with alternate fortune. This line of research aims at capturing, formally modeling and successfully exploiting all the *time-dependent characteristics* of the fundamental *big data model* ranging from *state model* to *query model*. Temporal big data management thus poses novel research challenges and exciting directions to be followed, and a first critical result is represented by recognizing that traditional time-focused models, techniques and algorithms developed in previous years are not suitable to deal with novel characteristics of big data, mainly due to *volume*, *heterogeneity* and *scalability issues*. Inspired by these considerations, in this paper we provide a comprehensive overview of state-of-the-art temporal big data management proposals, and criticisms on benefits and limitations of these initiatives. We complement our contributions with a deep discussion on future research directions in this area.**

***Keywords*—Big Data, Big Data Management, Temporal Aspects of Big Data Management**

I. Introduction

Big data management (e.g., [23], [2]) is recognized as one among the "hot-topics" in current database research. Among the well-known *3V properties* [36], i.e.: *volume*, *velocity* and *variety*, one *V* stands for *velocity* just to highlight that big data are inherently *temporal data*, for which *data streams* are some suitable instances (e.g., [6], [33], [29]). In line with this trend, a great deal of research efforts has been invested in *temporal aspects of big data management* during last years, with alternate fortune. Basically, this line of research aims at capturing, formally modeling and successfully exploiting all the *time-dependent characteristics* of the fundamental *big data model* (e.g., [32]) ranging from *query model* to *mining model*. This with relevant overlapping with previous experiences in *temporal database research* (e.g., [44]), which can be reasonably considered as the main conceptual background for this scientific area.

Indeed, currently a rich collection of research initiatives in the context of temporal aspects of big data management exists, ranging from *spatio-temporal modeling and characterization of big data* (e.g., [34]) to *high-performance sequential pattern mining from interval-based temporal (big) data* (e.g., [48]), from *change detection in temporally-evolving network (big) data* (e.g., [43]) to *temporal event tracing for big data analytics* (e.g., [38]), from *pattern mining from trajectory (big) data* (e.g., [60]) to *query optimization issues* (e.g., [35]), and so forth.

On the other hand, the need for *real-time processing of big data* (e.g., [39]), mostly driven by applicative requirements dictated by a wide range of real-life applications ranging from *big energy data analytics* (e.g., [46]) to *Cloud-based applications* (e.g., [57]), and from *sensor data management* (e.g., [58]) to *Web advertising* (e.g., [10]), and so forth, are leading the research scene, by originating a relatively large literature that also includes advanced topics like *high-performance* and *high-scalability* (e.g., [50], [30]).

Temporal big data management thus poses novel research challenges and exciting directions to be followed, and a first critical result is represented by recognizing that traditional time-focused models, techniques and algorithms developed in previous years are not suitable to deal with novel characteristics of big data, mainly due to *volume*, *heterogeneity* and *scalability issues*. The latter one represents an objective starting point from which moving future research efforts.

A leading instance of temporal big data is, without doubts, represented by (big) data streams (e.g., [39]), where the *time concept* plays a major role. For instance, it is enough to think of modern applications built on top of *energy (big) data streams* (e.g., [5]), *graph-like (big) data streams* (e.g., [8], [19]), Web (big) data streams (e.g., [10]), *tensor analysis* (e.g., [52], [45]), *bio-informatics* (e.g., [55], [7]) and so forth. It is a matter of fact, nevertheless, that data stream management research has captured the interest of wider and wider research communities since decades, with a major . Here, a lot of research problems arise, mainly concerning the issue of efficiently *querying* (e.g., [41]) and supporting *advanced analytics* (e.g., [11]) over such kind of data, thus defining a comprehensive literature that has dominated the research scene since several years.

A leading case study of temporal big data management is, instead, represented by so-called systems and tools for supporting *historical (big) data query answering* (e.g., [56], [54]). Here, the problem is that of providing answers to queries over big data that do not involve the actual (big) data but, rather, their *previous versions*, which are suitably indexed by means of *time-aware indexing data structures* (e.g., [28]). This paradigm represents the conceptual basis of emerging analytical engines based on *historical data analysis* (e.g., [61]).

Inspired by these considerations, in this paper we provide a comprehensive overview of state-of-the-art temporal big data management proposals, and criticisms on benefits and limitations of these initiatives. We complement our contributions

1530-1311/15 $31.00 © 2015 IEEE
DOI 10.1109/TIME.2015.31

with a deep discussion on future research directions in this area.

The remaining part of the paper is organized as follows. In Section II, we provide a comprehensive overview of state-of-the-art temporal big data management approaches, by highlighting benefits and limitations exposed by these trends. Section III contains open research challenges and future directions in this scientific field, aiming at achieving a significant milestone to be exploited by forthcoming research efforts. Finally, Section IV summarizes the main contributions of this research.

II. Temporal Aspects of Big Data Management: State-of-the-Art Analysis

In this Section, we focus the attention on state-of-the-art temporal big data management techniques, by highlighting their benefits and limitations.

[48] presents a framework for *supporting sequential pattern mining from large-scale, interval-based temporal data* without complete a-priori knowledge. The framework is called *PSEMiner*, and it also exposes high-performance thanks to parallel and quantitative features. PSEMiner introduces *a novel interactive temporal data mining algorithm* capable of extracting precise temporal properties of sequential patterns, and it shows to be extremely useful in order to detect unknown, anomalous and unexpected patterns.

[43] focuses the attention on the problem of *detecting changes in temporally-evolving computer networks* by means of big data models and paradigms, which confirms the eccentricity of such a technology. Here, authors aim at utilizing a big data framework in order to characterize the *behavior* of central nodes over time as to detect changes in large-evolving computer networks. The latter may indicate potential events such as cyber attacks, network failures or major shifts. Big data network streams are thus monitored in order to determine consistency and inconsistency over time from central nodes. The *MapReduce* [26] computational framework is exploited to this end.

[38] moves the attention to *an innovative patient-driven data architecture* that allows us to enable the rapid storing and flexible expansion of temporal big data, by founding on *NoSQL paradigms* [9]. In particular, the time-line mode is easily used to generate a visual representation of patient records, thus providing physicians with a solid reference for patient consultation. Data reformulation methods are applied in order to produce additional temporal and spatial data by providing MapReduce-based Cloud computing functionalities, and enhancing Data Mining methods over such data.

[60] considers the very interesting applicative setting represented by big trajectory data, a well-known case of temporal big data, and the issue of *retrieving moving object gathering patterns* from such data, being these patterns recognized as useful in order to predict anomalies in traffic systems. Authors argue that, in state-of-the-art proposals, discovering such patterns is a challenging problem due to the fact that large numbers of moving objects will generate large volumes of big trajectory data. The proposed method for dealing with this challenge is based on so-called *spatio-temporal graphs*, which model spatio-temporal changes that may happen among *clusters of moving objects*. Finally, a novel *gathering retrieving algorithm* is proposed by searching the maximal complete graphs which meet the input spatio-temporal constraints.

[35] investigates the fundamental problem of *supporting temporal query answering*, being this class of queries recognized as at the heart of a broad range of business and network intelligence applications ranging from consumer behavior analysis to trend analysis, from temporal pattern mining to sentiment analysis on social media, from cyber security to network monitoring, and so forth. In this context, authors propose a unified framework called *Fast Approximate Queryable* (FAQ). FAQ specially targets *temporal query processing on big data*, by making use of a novel composition of *data sketching*, wavelet-style differencing for *temporal compression*, and *quantization*. FAQ is capable of dealing with diverse kinds of query including distinct counts, set membership, frequency estimation, top-K, and so forth. Experiments proposed by authors clearly show the benefits deriving from FAQ in modern temporal big data settings.

[10] specifically focuses on *temporal analytics over big data for Web advertisement*, and proposes the novel framework *TiMR* that combines a time-oriented data processing systems with a MapReduce framework. In more details, TiMR considers well-known *Behavioral Targets* (BT) as the atomic unit of its reasoning, being these BT modeling *temporal-aware user Web interaction components* such as prior searches, page views, click logs, and so forth. In TiMR, users are allowed to perform analytics by means of temporal queries, which are ensured to be succinct, scale-out-agnostic, and easy to write. Finally, the efficiency of TiMR is improved thanks to suitable *cost-based query fragmentation* and *temporal partitioning schemes*.

[28] takes in consideration the issue of *constructing a multidimensional index for supporting distortion-free subsequence matching efficiently*. Authors recognize that, in previous research experiences, this task is very time-consuming since it generates a huge number of data sub-sequences in order to consider *all* possible positions and *all* possible query lengths. To deal with this challenge, authors propose a new methodology where each index construction step is optimized as to work well with very large time series databases. The idea is based on the so-called *store-and-reuse* concept, which introduces a dynamic programming technique capable of storing the intermediate results and reuse them in next steps repeatedly.

[58] focuses the attention on the problem of *detecting errors in big sensor data on Clouds*. According to the infrastructure argued by authors, big sensor data are processed on Cloud computing platforms (e.g., [16], [59]), which are capable of providing a flexible stack of massive computing, storage, and software services in a scalable manner, at a low cost. Nevertheless, this computational paradigm suffers from *severe data errors* that may occur in big sensor data. In order to deal with this drawback, authors propose a novel data error detection approach that exploits the full computation potential of Cloud platforms and the typical network features of *Wireless Sensor Networks* (WSN). Specifically, the proposed approach is based on the scale-free network topology and on the constraint predicating that most detection operations can be conducted in limited temporal or spatial data blocks instead

of a whole big data set, thus accelerating the detection and location process dramatically.

[49] is a comprehensive study on the problem of *mining and forecasting big time series data*, whose processing is relevant for a wide family of applications ranging from large-scale tensor analysis to non-linear modeling and automatic mining. In particular, authors provide an overview on the following key topics: (*i*) similarity search and pattern discovery; (*ii*) linear modeling and summarization; (*iii*) non-linear modeling and forecasting; (*iv*) time-series mining and tensor analysis.

[3] considers the issue of *supporting efficient storage of GPS (temporal) big data*. In the proposed approach, *data are partitioned into cells giving priority to geo-spatial location*. The geography of an area is divided into suitable data cells, and the size of such data cells is determined based on the previously-observed location-specific queries on that area. This criterion enables computation to occur closer to data location. *Data redundancy* is also embedded as to enable fail-over mechanisms and to improve performance.

[5] moves the attention on *big energy data*, and studies the problem of *supporting real-time integration* of these data, which are recognized as a classical instance of big data. The basic idea proposed in this paper consists in developing a suitable *Energy Management System* (EMS) based on a *Data Stream Management System* (DSMS) (e.g., [1]) rather than conventional DBMS, like in traditional approaches. To this end, several extensions of data stream processing devoted to support *real-time features* are proposed and experimentally assessed.

[51] explores challenges and opportunities of *mobile Complex Event Processing* (CEP) for supporting *real-time big data applications*. Authors correctly state that the expansion of mobile applications in various domains has proposed the issue of the optimal usage of limited resources that characterize such mobile devices (e.g., limited bandwidth, limited data storage availability, limited processing power), with particular regards to the case of real-time big data applications. Many of such applications are indeed oriented towards a natural synergy with event processing methodologies, as particular situations may cause the occurrence of a particular computational setting or another. In line with this trend, authors argue that real-time, big-data-driven applications require a novel infrastructure for distributed CPE that is only *partially* executed on mobile devices. A novel architecture that founds on these predicates is thus presented and discussed in details.

[4] considers the issue of supporting *real-time QoS monitoring for Cloud-based big data analytics applications in mobile environments*. The proposed framework, called CLAMS (*Cross-Layer Multi-Cloud Applicaion Monitoring-as-a-Service Framework*), is designed for providing essential information for effective and efficient big data analytics, and it performs the following tasks: (*i*) multi-cloud monitoring; (*ii*) cross-layer monitoring of applications. The framework is finally implemented on multi-cloud platforms integrating Amazon and Azure.

[53] describes the anatomy of *an architecture for detecting events in real-time using massive heterogeneous data sources*. The proposed architecture focuses on the problem of monitoring and reacting to all sort of circumstances that may occur in massive data sources, from network congestion to natural catastrophes. The main problem is thus that of *processing big data in real-time*, in a proactive manner. As a consequence, it becomes mandatory to develop a platform for aggregating all the necessary information and orchestrating it in the best-possible way, as to meet the application requirements above. Indeed, sources to be integrated are noticeable in number and different in nature: from sensors to smart phones, from GPS signals to social networks, and so forth. To cope with these challenges, authors propose an architecture for gathering, aggregating and processing a wide range of sensory input coming from very heterogeneous data sources, which also includes *the active involvement of citizens*.

[40] focuses on big trajectory data generated by GPS-equipped mobile devices, and studies *a novel path finding query class* that finds the *most frequent path* (MFP) during user-specified time periods in large-scale historical trajectory data. Authors recognize that, while there exist in literature several proposals relying on the MFP concept, all consider a *fixed* time period, and, in addition to this, they do not consider the following three key properties: (*i*) *suffix-optimality*, (*ii*) *length-insensitivity*, (*iii*) *bottleneck-free*. In order to fulfill this gap, they propose a composite framework that both supports the *flexible* MFP time period and the three desired properties, plus a very extensive experimental assessment and analysis of this framework against real-life big trajectory data sets.

[42] provides a very interesting case study illustrating the challenges of *effectively and efficiently supporting real-time query suggestion on Twitter data*, the leading big social data repository at now. In more details, the solution proposed by authors has two versions. The first one, is implemented on top of a typical Hadoop-based analytics stack. The second one, proposed in order to meet the *latency requirements* necessary to generate meaningful real-time results, is represented by *a custom in-memory processing engine* specifically developed for the supported task. The lesson learned from this research effort is that wile Hadoop is still the most popular big data platform today, it is not well-suited to low-latency processing, thus opening the door to the issue of effectively and efficiently processing not only big but also *fast* data.

[47] proposes a nice work on *mining trillions of time series sub-sequences under Dynamic Time Warping* (DTW), by overcoming the state-of-the-art *Euclidean-distance-based search algorithms*. The framework proposed by authors is very large-scale, and it clearly represents a real-life instance of temporal big data management system, being also targeted to industry and science application scenarios that typically explore billions of time series objects for applicative and analysis goals. Authors prove the effectiveness and efficiency of their research via a rich series of experiments, whose results clearly overcome state-of-the-art achievements.

Finally, [27] focuses the attention on the problem of *finding relevant and appropriately-ranked information at the right time on the Social Web*. Basically, authors propose *real-time collaborative filtering algorithms* for big fast data and apply them to the relevant application scenario represented by the Twitter social network.

III. TEMPORAL ASPECTS OF BIG DATA MANAGEMENT: FUTURE RESEARCH DIRECTIONS

In this Section, we discuss most relevant open research challenges and future directions that concern the development and the advancement of temporal big data management techniques.

Modeling Temporal Big Data While the time concept and the temporal dimension are quite explicit in temporal big data, the issue of *effectively and flexibly capturing and modeling time in big data* still demands for additional research efforts, for example oriented to describing complex instances like *interval-based temporal big data sets*.

Indexing Temporal Big Data Designing *cost-optimal indexing data structures* based on the temporal dimension of big data is still an open issue, mainly due to the enormous size of big data. This poses several challenges to deal with, cause many correlated problems arise, such as the issue of dynamically maintaining and updating these indexing structures.

In-Memory Processing Engines for Temporal Big Data Management Dealing with temporal big data can be *partially* performed by means of *highly-efficient in-memory processing engines*, which may act within the broader computational setting represented by a distributed, heterogeneous architecture that also includes Hadoop-based components. These *hybrid solutions* are among the next-generation architecture for processing temporal big data, but the integration between in-memory engines and classical-big-data modules is still and open problem to be further investigated.

Storage Issues of Temporal Big Data Management How to effectively and efficiently represent massive temporal big data sets in *secondary memory*? This question put emphasis on the challenge predicating that the presence of the time concept and the temporal dimension in big data conveys in *data representation challenges* due to the fact that, for instance, big data analytics systems would require to represent in secondary memory different *versions* (over time) of the target temporal big data set, thus involving in an *exponential explosion* of the final disk occupancy (along with data access drawbacks). Issues like those mentioned above need meticulous attention and research efforts during next years.

Scalability Issue of Temporal Big Data Management Big data expose, by definition, high-volumes. When the temporal dimension is explicit and prominent in big data sets, *scalability issues* become more problematic, due to the need of apply mining procedures that, following the natural *hierarchical organization* of the temporal dimension, must access and process data repetitively. This poses severe scalability challenges, perhaps tamed by Hadoop-based temporal big data management architectures.

Approximate Query Answering Techniques over Temporal Big Data Temporal big data are intrinsically large-scale and enormous-in-size. Given these peculiarities, *approximate query answering*, which has already proved to be successful on very-large data sets (e.g., [18], [17]), is a promising research line to be considered in future.

Ranking Temporal Big Data *Ranking* is a common task in many data-intensive systems. How to rank temporal big data? Which features should be considered during the ranking of temporal big data? Nowadays, very pioneering approaches on this critical topic exist, while it turns to be very relevant for a wide spectrum of data-intensive primitives implemented within the core layers of big data analytics systems.

Imprecise and Uncertain Temporal Big Data Management Due to the same nature of the (industrial and scientific) processes that generate them, temporal big data sets are intrinsically *imprecise* and *uncertain* (e.g., [12], [37], [20]). This evident property asks for novel models, techniques and algorithms for processing (imprecise and uncertain) temporal big data, beyond the classical solutions that are, instead, developed under the assumption of dealing with (exact) big data, which is poorly supported in real-life big data applications and systems.

Innovative Analytical Queries Temporal big data management is a genuine task within the main big data analytics process (e.g., [24], [13], [25]). In this context, devising *innovative analytical queries* that retrieve time-aware answers is a first-class challenge for next-generation big data research.

Advanced Mining Predicates on Temporal Big Data Temporal big data are naturally prone to be mined by means of a wide family of techniques and algorithms, such as *frequent itemset mining*, *outlier detection*, and so forth. Nevertheless, due to the particular time-oriented nature of such data, finer-grain advanced mining predicates, such as *multi-granularity analysis*, *mining over multiple time-overlapping big data*, *apriori time-aware mining* and so forth, are necessary. All these mining predicates are, in turn, based on a rich collection of temporal big data management tasks, which are the main target to be pointed-out by future research efforts.

Multidimensional Analysis of Temporal Big Data Within the wider scope of analytics over temporal big data, *multi-dimensional analysis of temporal big data* can be obtained by applying well-consolidated *OLAP methodologies* [31] to such data, where the temporal dimension is indeed prominent (e.g., [15]). The latter is still an open issue in actual research.

Secure Access to Temporal Big Data Sets The issue of *supporting secure access to temporal big data sets* is of major relevance at now. Indeed, the explicit presence of the time concept and the temporal dimension in big data opens the door to novel challenges to be faced-off. For instance, access grants to big data may depend on their *temporal characterization*, following specific events that may occur in the target business organization. More generally, the temporal dimension adds more problematic challenges to the issue of securely access (temporal) big data, with novel perspectives of research with respect to traditional approaches.

Privacy-Preserving Temporal Big Data Management A relevant problem concerning the management of temporal big data concerns with the issue of supporting *privacy-preserving features*, i.e. preserving the privacy of (temporal) big data during the execution of typical data management tasks (e.g., accessing, indexing, query answering, and so forth). The relevance of this research challenge is becoming higher and higher (e.g., [21], [22], [14]), even due to modern Cloud infrastructures where machines need to exchange bulks of data very frequently, while the temporal notation makes this issue harder and harder.

Temporal Big Data Visualization Tools *Visualizing big data* is a well-known problem that puts its roots in a very wide family of industrial and scientific application scenarios. When dealing with temporal big data, the need for rich and flexible visualization tools is evident, since the temporal dimension is naturally prone to be analyzed and explored according to *different levels of granularity and resolution*, thus enriching the knowledge fruition phase from (temporal) big data.

Integration with NoSQL Platforms NoSQL platforms seem to be the most suitable computational platforms devoted to temporal big data management. Nevertheless, the integration of models, techniques and algorithms for processing temporal big data starting from their original format, mainly born in the context of industrial and scientific applications, with emerging NoSQL platform is still an open problem that requires much effort during future years.

IV. CONCLUSIONS

In this paper, we provided a comprehensive overview of state-of-the-art temporal big data management techniques, by highlighting their benefits and limitations, and a wide-ranging discussion on open research challenges and future directions in this scientific field, aiming at achieving a significant milestone to be exploited by forthcoming research efforts.

REFERENCES

[1] D. J. Abadi, D. Carney, U. Çetintemel, M. Cherniack, C. Convey, C. Erwin, E. F. Galvez, M. Hatoun, A. Maskey, A. Rasin, A. Singer, M. Stonebraker, N. Tatbul, Y. Xing, R. Yan, and S. B. Zdonik. Aurora: A data stream management system. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, San Diego, California, USA, June 9-12, 2003*, page 666, 2003.

[2] D. Agrawal, S. Das, and A. El Abbadi. Big data and cloud computing: current state and future opportunities. In *EDBT*, pages 530–533, 2011.

[3] P. K. Akulakrishna, J. Lakshmi, and S. K. Nandy. Efficient storage of big-data for real-time GPS applications. In *2014 IEEE Fourth International Conference on Big Data and Cloud Computing, BDCloud 2014, Sydney, Australia, December 3-5, 2014*, pages 1–8, 2014.

[4] K. Alhamazani, R. Ranjan, P. P. Jayaraman, K. Mitra, M. Wang, Z. G. Huang, L. Wang, and F. A. Rabhi. Real-time qos monitoring for cloud-based big data analytics applications in mobile environments. In *IEEE 15th International Conference on Mobile Data Management, MDM 2014, Brisbane, Australia, July 14-18, 2014 - Volume 1*, pages 337–340, 2014.

[5] D. Anjos, P. Carreira, and A. P. Francisco. Real-time integration of building energy data. In *2014 IEEE International Congress on Big Data, Anchorage, AK, USA, June 27 - July 2, 2014*, pages 250–257, 2014.

[6] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom. Models and issues in data stream systems. In *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 3-5, Madison, Wisconsin, USA*, pages 1–16, 2002.

[7] S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. E. Johnson, and P. J. O'Connor. Data mining for censored time-to-event data: a bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Min. Knowl. Discov.*, 29(4):1033–1069, 2015.

[8] J. J. Cameron, A. Cuzzocrea, F. Jiang, and C. K. Leung. Frequent pattern mining from dense graph streams. In *Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014), Athens, Greece, March 28, 2014.*, pages 240–247, 2014.

[9] R. Cattell. Scalable SQL and nosql data stores. *SIGMOD Record*, 39(4):12–27, 2010.

[10] B. Chandramouli, J. Goldstein, and S. Duan. Temporal analytics on big data for web advertising. In *IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012*, pages 90–101, 2012.

[11] A. Cuzzocrea. CAMS: olaping multidimensional data streams efficiently. In *Data Warehousing and Knowledge Discovery, 11th International Conference, DaWaK 2009, Linz, Austria, August 31 - September 2, 2009, Proceedings*, pages 48–62, 2009.

[12] A. Cuzzocrea. Retrieving accurate estimates to OLAP queries over uncertain and imprecise multidimensional data streams. In *Scientific and Statistical Database Management - 23rd International Conference, SSDBM 2011, Portland, OR, USA, July 20-22, 2011. Proceedings*, pages 575–576, 2011.

[13] A. Cuzzocrea. Analytics over big data: Exploring the convergence of datawarehousing, OLAP and data-intensive cloud infrastructures. In *37th Annual IEEE Computer Software and Applications Conference, COMPSAC 2013, Kyoto, Japan, July 22-26, 2013*, pages 481–483, 2013.

[14] A. Cuzzocrea. Privacy and security of big data: Current challenges and future research perspectives. In *Proceedings of the First International Workshop on Privacy and Secuirty of Big Data, PSBD@CIKM 2014, Shanghai, China, November 7, 2014*, pages 45–47, 2014.

[15] A. Cuzzocrea, L. Bellatreche, and I. Song. Data warehousing and OLAP over big data: current challenges and future research directions. In *Proceedings of the sixteenth international workshop on Data warehousing and OLAP, DOLAP 2013, San Francisco, CA, USA, October 28, 2013*, pages 67–70, 2013.

[16] A. Cuzzocrea, G. Fortino, and O. F. Rana. Managing data and processes in cloud-enabled large-scale sensor networks: State-of-the-art and future research directions. In *13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2013, Delft, Netherlands, May 13-16, 2013*, pages 583–588, 2013.

[17] A. Cuzzocrea, F. Furfaro, S. Greco, E. Masciari, G. M. Mazzeo, and D. Saccà. A distributed system for answering range queries on sensor network data. In *3rd IEEE Conference on Pervasive Computing and Communications Workshops (PerCom 2005 Workshops), 8-12 March 2005, Kauai Island, HI, USA*, pages 369–373, 2005.

[18] A. Cuzzocrea, F. Furfaro, E. Masciari, D. Saccà, and C. Sirangelo. Approximate query answering on sensor network data streams. In *GeoSensor Networks*, pages 53–72. 2004.

[19] A. Cuzzocrea, F. Jiang, and C. K. Leung. Frequent subgraph mining from streams of linked graph structured data. In *Proceedings of the Workshops of the EDBT/ICDT 2015 Joint Conference (EDBT/ICDT), Brussels, Belgium, March 27th, 2015.*, pages 237–244, 2015.

[20] A. Cuzzocrea, C. K. Leung, and R. K. MacKinnon. Mining constrained frequent itemsets from distributed uncertain data. *Future Generation Comp. Syst.*, 37:117–126, 2014.

[21] A. Cuzzocrea, V. Russo, and D. Saccà. A robust sampling-based framework for privacy preserving OLAP. In *Data Warehousing and Knowledge Discovery, 10th International Conference, DaWaK 2008, Turin, Italy, September 2-5, 2008, Proceedings*, pages 97–114, 2008.

[22] A. Cuzzocrea and D. Saccà. Balancing accuracy and privacy of OLAP aggregations on data cubes. In *DOLAP 2010, ACM 13th International Workshop on Data Warehousing and OLAP, Toronto, Ontario, Canada, October 30, 2010, Proceedings*, pages 93–98, 2010.

[23] A. Cuzzocrea, D. Saccà, and J. D. Ullman. Big data: a research agenda. In *17th International Database Engineering & Applications Symposium, IDEAS '13, Barcelona, Spain - October 09 - 11, 2013*, pages 198–203, 2013.

[24] A. Cuzzocrea, I. Song, and K. C. Davis. Analytics over large-scale multidimensional data: the big data revolution! In *DOLAP 2011, ACM 14th International Workshop on Data Warehousing and OLAP, Glasgow, United Kingdom, October 28, 2011, Proceedings*, pages 101–104, 2011.

[25] A. Cuzzocrea and I.-Y. Song. Big graph analytics: The state of the art and future research agenda. In *Proceedings of the 17th International Workshop on Data Warehousing and OLAP, DOLAP 2014, Shanghai, China, November 3-7, 2014*, pages 99–101, 2014.

[26] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, 2008.

[27] E. Diaz-Aviles, W. Nejdl, L. Drumond, and L. Schmidt-Thieme. Towards real-time collaborative filtering for big fast data. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, pages 779–780, 2013.

[28] M. Gil, B. Kim, M. Choi, and Y. Moon. Fast index construction for distortion-free subsequence matching in time-series databases. In *2015 International Conference on Big Data and Smart Computing, BIGCOMP 2015, Jeju, South Korea, February 9-11, 2015*, pages 130–135, 2015.

[29] L. Golab and M. T. Özsu. Issues in data stream management. *SIGMOD Record*, 32(2):5–14, 2003.

[30] T. Goldschmidt, A. Jansen, H. Koziolek, J. Doppelhamer, and H. P. Breivold. Scalability and robustness of time-series databases for cloud-native monitoring of industrial processes. In *2014 IEEE 7th International Conference on Cloud Computing, Anchorage, AK, USA, June 27 - July 2, 2014*, pages 602–609, 2014.

[31] J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub totals. *Data Min. Knowl. Discov.*, 1(1):29–53, 1997.

[32] H. Gregersen and C. S. Jensen. Temporal entity-relationship models - A survey. *IEEE Trans. Knowl. Data Eng.*, 11(3):464–497, 1999.

[33] G. Hulten, L. Spencer, and P. M. Domingos. Mining time-changing data streams. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, August 26-29, 2001*, pages 97–106, 2001.

[34] H. Jiang, L. Huang, J. J. Zhang, Y. Zhang, and D. W. Gao. Spatial-temporal characterization of synchrophasor measurement systems - A big data approach for smart grid system situational awareness. In *48th Asilomar Conference on Signals, Systems and Computers, ACSSC 2014, Pacific Grove, CA, USA, November 2-5, 2014*, pages 750–754, 2014.

[35] U. Khurana, S. Parthasarathy, and D. S. Turaga. FAQ: A framework for fast approximate query processing on temporal data. In *Proceedings of the 3rd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, BigMine 2014, New York City, USA, August 24, 2014*, pages 29–45, 2014.

[36] D. Laney. 3D data management: Controlling data volume, velocity, and variety. Technical report, META Group, 2001.

[37] C. K. Leung, A. Cuzzocrea, and F. Jiang. Discovering frequent patterns from uncertain data streams with time-fading and landmark models. *T. Large-Scale Data- and Knowledge-Centered Systems*, 8:174–196, 2013.

[38] C. Lin, L. Huang, S. T. Chou, C. Liu, H. Cheng, and I. Chiang. Temporal event tracing on big healthcare data analytics. In *2014 IEEE International Congress on Big Data, Anchorage, AK, USA, June 27 - July 2, 2014*, pages 281–287, 2014.

[39] X. Liu, N. Iftikhar, and X. Xie. Survey of real-time processing systems for big data. In *18th International Database Engineering & Applications Symposium, IDEAS 2014, Porto, Portugal, July 7-9, 2014*, pages 356–361, 2014.

[40] W. Luo, H. Tan, L. Chen, and L. M. Ni. Finding time period-based most frequent path in big trajectory data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2013, New York, NY, USA, June 22-27, 2013*, pages 713–724, 2013.

[41] G. S. Manku and R. Motwani. Approximate frequency counts over data streams. *PVLDB*, 5(12):1699, 2012.

[42] G. Mishne, J. Dalton, Z. Li, A. Sharma, and J. Lin. Fast data in the era of big data: Twitter's real-time related query suggestion architecture. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2013, New York, NY, USA, June 22-27, 2013*, pages 1147–1158, 2013.

[43] J. M. Namayanja and V. P. Janeja. Change detection in temporally evolving computer networks: A big data framework. In *2014 IEEE International Conference on Big Data, Big Data 2014, Washington, DC, USA, October 27-30, 2014*, pages 54–61, 2014.

[44] G. Özsoyoglu and R. T. Snodgrass. Temporal and real-time databases: A survey. *IEEE Trans. Knowl. Data Eng.*, 7(4):513–532, 1995.

[45] E. E. Papalexakis, U. Kang, C. Faloutsos, N. D. Sidiropoulos, and A. Harpale. Large scale tensor decompositions: Algorithmic developments and applications. *IEEE Data Eng. Bull.*, 36(3):59–66, 2013.

[46] J. Peral, A. Ferrández, R. Tardío, A. Maté, and E. de Gregorio. Energy consumption prediction by using an integrated multidimensional modeling approach and data mining techniques with big data. In *Advances in Conceptual Modeling - ER 2014 Workshops, ENMO, MoBiD, MReBA, QMMQ, SeCoGIS, WISM, and ER Demos, Atlanta, GA, USA, October 27-29, 2014. Proceedings*, pages 45–54, 2014.

[47] T. Rakthanmanon, B. J. L. Campana, A. Mueen, G. E. A. P. A. Batista, M. B. Westover, Q. Zhu, J. Zakaria, and E. J. Keogh. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *TKDD*, 7(3):10, 2013.

[48] G. Ruan, H. Zhang, and B. Plale. Parallel and quantitative sequential pattern mining for large-scale interval-based temporal data. In *2014 IEEE International Conference on Big Data, Big Data 2014, Washington, DC, USA, October 27-30, 2014*, pages 32–39, 2014.

[49] Y. Sakurai, Y. Matsubara, and C. Faloutsos. Mining and forecasting of big time-series data. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 919–922, 2015.

[50] S. Shahrivari. Beyond batch processing: Towards real-time and streaming big data. *Computers*, 3(4):117–129, 2014.

[51] N. Stojanovic, L. Stojanovic, Y. Xu, and B. S. Nissatech. Mobile CEP in real-time big data processing: challenges and opportunities. In *The 8th ACM International Conference on Distributed Event-Based Systems, DEBS '14, Mumbai, India, May 26-29, 2014*, pages 256–265, 2014.

[52] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos. Incremental tensor analysis: Theory and applications. *TKDD*, 2(3), 2008.

[53] G. Valkanas, D. Gunopulos, I. Boutsis, and V. Kalogeraki. An architecture for detecting events in real-time using massive heterogeneous data sources. In *Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, BigMine 2013, Chicago, IL, USA, August 11, 2013*, pages 103–109, 2013.

[54] F. Wang and C. Zaniolo. Temporal queries and version management in xml-based document archives. *Data Knowl. Eng.*, 65(2):304–324, 2008.

[55] H. Wang, H. Sun, T. Chang, H. Lo, W. Cheng, G. C. Tseng, C. Lin, S. Chang, N. R. Pal, and I. Chung. Discovering monotonic stemness marker genes from time-series stem cell microarray data. *BMC Genomics*, 16(S-2):S2, 2015.

[56] W. H. Wang, R. Liu, D. Theodoratos, and X. Wu. Efficient storage and temporal query evaluation in hierarchical data archiving systems. In *Scientific and Statistical Database Management - 23rd International Conference, SSDBM 2011, Portland, OR, USA, July 20-22, 2011. Proceedings*, pages 109–128, 2011.

[57] M. Westerlund, U. Hedlund, G. Pulkkis, and K. Björk. A generalized scalable software architecture for analyzing temporally structured big data in the cloud. In *New Perspectives in Information Systems and Technologies, Volume 1 [WorldCIST'14, Madeira Island, Portugal, April 15-18, 2014]*, pages 559–569, 2014.

[58] C. Yang, C. Liu, X. Zhang, S. Nepal, and J. Chen. A time efficient approach for detecting errors in big sensor data on cloud. *IEEE Trans. Parallel Distrib. Syst.*, 26(2):329–339, 2015.

[59] B. Yu, A. Cuzzocrea, D. H. Jeong, and S. Maydebura. On managing very large sensor-network data using bigtable. In *12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2012, Ottawa, Canada, May 13-16, 2012*, pages 918–922, 2012.

[60] J. Zhang, J. Li, S. Wang, Z. Liu, Q. Yuan, and F. Yang. On retrieving moving objects gathering patterns from trajectory data via spatio-temporal graph. In *2014 IEEE International Congress on Big Data, Anchorage, AK, USA, June 27 - July 2, 2014*, pages 390–397, 2014.

[61] J. Zheng and A. Dagnino. An initial study of predictive machine learning analytics on large volumes of historical data for power system applications. In *2014 IEEE International Conference on Big Data, Big Data 2014, Washington, DC, USA, October 27-30, 2014*, pages 952–959, 2014.