Available online at https://www.asjp.cerist.dz/en/PresentationRevue/134

RIST

Information Processing at the Digital Age Journal

Volume 26, No 1 (2021)

Applications of Graph Databases and Big Data Technologies in Healthcare

Faiza Deghmani[a,*], Idir AmineAmarouche[a], Kamel Boukhalfa[a]

[a]LSI, Department of Computer Science, University of Science and Technology Houari Boumediene (USTHB)
BP 32 El Alia 16111 Bab Azzouar Algiers, Algeria

Abstract

Several aspects related to big data technologies in the healthcare area, like architecture and capabilities, have been surveyed. Also, many works propose the use of graph databases in healthcare domain. However, according to the best of our knowledge, there is no work that addresses the challenges related to big data technologies and graph databases in healthcare. For this reason, we address a survey of big data in healthcare based on a graph database. The presented paper exposes a gap analysis based on a set of paper related to the healthcare systems based on graph databases and big data technologies.

Keywords: Big Data, Big Data Technologies, NoSql Databases, Graph Databases, Healthcare

1. Introduction

In recent years, digitized data in healthcare are generated at very high speed with the data coming in from internal as well as external sources, such as, mobile devices, wearable sensor devices, Electronic Health Records (EHR), social media and remote health monitoring devices (Mathew et Pillai 2015). These data have a big volume and a variety of formats (Images, texts, photos...etc.) and a high level of velocity. As a result, these data meet the main characteristics of big data and motivate the need for required management solutions , see Zillner and Neururer, 2016; Mathew and Pillai 2015.

[*]Tel: +213 772 373 621, E-mail address: deghmani.faiza@gmail.com

Furthermore, rising rates of chronic diseases, increased population, need for evidence-based medicine, inability to process and get insight from ever-increasing heterogeneous health data are also drivers for adopting big data solutions in healthcare field, Mathew and Pillai, 2015. Big data technologies should provide capabilities to manage massive data available in the healthcare industry which need to work on prediction, prevention and personalization to improve their outcomes and the quality of patient care, Mathew and Pillai, 2015.

This trend can be explained by the limits encountered by relational databases, which are not designed to effectively cope with such large quantities of data and promote the development of NoSQL databases, Fraczek and Plechawska-Wojcik, 2017. Several works, show that NoSQL databases provide significant advantages, such as, easy and automatic scaling, better performance and high availability which address the limitations of relational databases in distributed healthcare systems, Ercan and Lane, 2014. According to Mathew and Pillai, 2015, NoSQL databases expose four data store classes:(1) Key Value oriented database which presents data stored as couples of values and their keys; (2) Column oriented database which stores data as columns and each column has a key; (3) Document oriented database, in which data are stored in documents (JSON format) and (4) Graph oriented database in which data are represented as a network and stored as nodes and edges.

Recently, graph databases regained an important interest among the researchers for many reasons. The inherent property of graphs as a structure is that represents the strong connectivity within the data. Graph databases are the best for dealing with complex, semi-structure, and mainly densely connected data and it is very fast in terms of queries, Kaliyar, 2015. The use of graph databases in healthcare has significant benefits, Park et al., 2014. That's why many researchers proposed the use of graph databases in healthcare systems to offer better analytics either descriptive or predictive, Ling et al., 2014; Khan et al., 2016; Sen et al., 2017. Also, to understand relationships between entities and to construct efficient data management framework for large scale healthcare system Park et al., 2014; Khan et al, 2016. In addition, there are several works that survey many challenges related to big data technologies in medical and health area like architecture and capabilities, Wang and Hajli, 2017; Zillner and Neururer, 2016; Krishnan, 2016; Mathew and Pillai, 2015; Asare-Frempong and Jayabalan, 2017; Wang et al., 2015; Raghupathi and Raghupathi, 2014; Cyganek et al., 2015; Wang and Hajli, 2017. The use of NoSQL databases in the health sector was also evaluated, Yaqoob et al., 2016. But to the best of our knowledge, there is no work which has addressed a survey of big data technologies in healthcare based on graph databases. To address this lack, the present paper proposes to review recent papers related to healthcare systems based on graph databases and big data technologies.

This paper is organized as follows. The next section presents background concepts related to graph databases, big data technologies and healthcare data. Section 3 motivates the present paper. Section 4 presents the reviewed work that deal with adoption of graph databases and big data technologies to handle healthcare system requirement. In section 5, we give an analysis of studied systems presented in the previous section and highlight potential challenges. Finally, section 6 concludes the paper and presents our perspectives.

2. Background

In this section, we provide the main definitions related to the core concepts used in the proposed survey.

2.1. Big data

The term big data refers to the huge amount of data that needs new technologies and architectures to find valuable knowledge from it by using new and innovative analysis practices. Various explanations from 3V (i.e. Volume, Variety, and Velocity) to 4V (i.e. Volume, Velocity, Variety and Veracity) have been provided to define big data(Yaqoob et al. 2016). The term "volume" refers to the size of the data, "velocity" refers to the speed of incoming and outgoing data, and “variety” describes the sources and types of data. “Veracity” or “variability” as the fourth V; refers to the messiness and trustworthiness of data, Yaqoob et al., 2016. Big data is also defined by “Value” which refers to the worth of hidden insights inside big data.

To identify the potential benefits offered by big data, it is necessary to understand its architecture and component functionalities. As illustrated by the figure 1, big data architecture consists of a data-based logical framework that starts with data capture, proceeds via data transformation, and concludes with data consumption. According to Wang et al., 2015, this architecture is composed of five major layers:

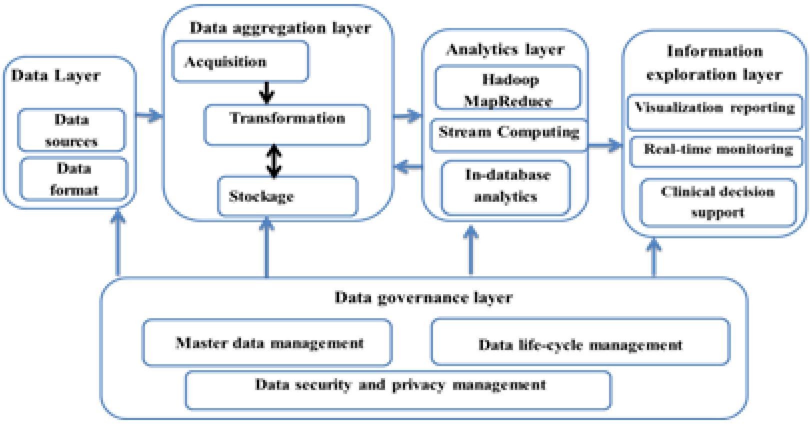

Fig. 1. Big data Architecture

Data layer: This layer focus on data sources and content format. The data is divided into structured data, semi-structured data, and unstructured data. These data are collected from various locations, and will be stored immediately into appropriate databases, depending on the content format.

Data aggregation layer: It is responsible for handling data. Data will be processed by performing three steps: data acquisition, transformation, and storage. The goal of data acquisition is to read collected data. The transformation engine must be capable of moving, cleaning, splitting, translating, merging, sorting, and validating data. Finally, the data are loaded into the target databases such as Hadoop Distributed File Systems (HDFS) or in a Hadoop cloud for further processing and analysis.

Analytics Layer: This layer is responsible for processing all kinds of data and performing appropriate analyses. Data analysis can be divided into three components: Hadoop Map/Reduce, stream computing, and in-database analytics, depending on the type of data and the purpose of the analysis.

Information exploration layer: It generates outputs such as the various visualization reporting options, real-time monitoring of information, and meaningful business insights derived from the analytics platforms to users in the organization.

Big data governance layer: This layer is composed of Master Data Management (MDM), data life-cycle management, and data security and privacy management that emphasize how to harness data in the organization. The MDM is regarded as the processes, governance, policies, standards, and tools for managing data. Data is properly standardized, removed, and incorporated in order to create the immediacy, completeness, accuracy, and availability of master data for supporting data analysis. The data life-cycle management is the process of managing business information throughout its lifecycle, from archiving data, through maintaining data warehouse, testing and delivering different application systems to deleting and disposing of data. Data security and privacy management is the platform for providing enterprise-level data activities in terms of discovery, configuration assessment, monitoring, auditing, and protection.

2.2. Healthcare data

Healthcare information systems, social media and medical devices are some of the main providers of health data according to Zillner and Neururer, 2016; Asare-Frempong and Jayabalan, 2017. Healthcare social websites, such as “PatientsLikeMe†” are generating large sets of health data, by voluntarily sharing data about rare diseases or remarkable experiences with common diseases, Zillner and Neururer, 2016. Sensors such as glucose monitors or blood pressure monitors can provide some valuable insights about patients’ health conditions, Asare-Frempong and Jayabalan, 2017. Data provided by hospital information system such as patients’ demographics, medical records, lab results and medical images, cost and billing data constitute the Electronic Medical Record (EMR). EMR is defined as the record of the periodic care provided mainly by one institution, Ercan and Lane, 2014. EHR is defined as an electronic record that holds a patient’s lifetime health-related information or a collection of EMRs for a single individual, Ercan and Lane, 2014. Consequently, health data are distributed, heterogeneous in terms of structure, feature and semantic which makes them particularly challenging to secure, to store, to process, to share and to analyze.

2.3. Big data in healthcare

Several developments in healthcare sector, such as escalating healthcare costs, increased need for healthcare coverage, and shifts in provider reimbursement trends, trigger the demand for big data technologies in order to improve the overall efficiency and quality of care delivery, see Zillner andt Neururer, 2016. Accordingly, taking into consideration its complexity, heterogeneity, fast growing and size we need special tools to analyse it and we should consider healthcare data as big data. Such a situation imparts to healthcare data a 5Vs character ,Cyganek et al., 2015, namely:

Volume: When doctor’s note stored as a text file is a few kilo bytes; a raw image requires a few megabytes and sophisticated diagnostic tools such as MRI‡ requires giga bytes. If such a volume size is multiplied by the number of test carried out in the hospital, we should be ready to deal with tera and peta bytes. According to Asare-Frempong and Jayabalan, 2017, the data amassed in the healthcare industry is about 500 petabytes by estimate in 2012.

Velocity: Health data is in motion; new information about patient is added, and some medical records are updated. Therefore, some smart analytic tools applied to EHR analysis require that the models will not be rebuilt from scratch when new data come, but it will be improved. The need to process data in real time coming from streaming data like Remote Patient Monitoring, data from sensor devices and Telemedicine, Mathew and Pillai, 2015.

Variety: The EHR include data with heterogeneous structures, i.e., on the one hand structured data in the form of standardized medical information, such as DICOM§, or using the ICD** codes, but on the other hand the most valuable data could be found in unstructured data such as doctor's notes written in natural language, see Cyganek et al., 2015.

---

† http://www.patientslikeme.com/
‡ Magnetic Resonance Imaging
§ Digital imaging and communications in medicine
** International Classification of Diseases

Veracity: Data in the healthcare may be noisy and biased, and we may find abnormality as outliers. In addition, human errors are as important issue as well, due to the clinical mistakes and it consequences.

Value: Big data analytics tool deployment makes sense if it leads to healthcare improvement. Big data adoption in healthcare was not only to manage the massive health data, analytics of big data can be also applied in the diagnosis of diseases and in the treatment of illnesses, which makes the applications of big data analytics in healthcare a solution to improve the quality of care and allow advancing research in healthcare Zillner and Neururer, 2016; Krishnan, 2016.

Big Data technologies will definitely open new opportunities and enable breakthroughs related to, among the others healthcare data analytics addressing different perspectives: (i) descriptive to answer what happened, (ii) diagnostic to answer the reason why it happened, (iii) predictive to understand what will happen and (iv) prescriptive to detect how we can make it happen, Heinrich et al., 2016.

2.4 Big data analytics capabilities in healthcare context

The logical layers presented above would enable healthcare managers to understand how to transform the healthcare data from various sources into meaningful clinical information through big data implementations. In this context, big data analytics capability is defined as the ability to acquire, store, process and analyze large amounts of health data in various forms, and deliver meaningful information to users, Wang and Hajli, 2017. These capabilities are derived from the various design principles and functionalities of big data and are confirmed by the real-world use of big data in healthcare contexts. According to Wang et al., 2015, the main capabilities are described as follows:

Traceability: Traceability is the ability to track output data from all the system's IT components throughout the healthcare's setting units. Thus, big data can track information that is created by the medical devices in real time. This makes it possible to gather location, event and physiological information from each patient wearing the device. This information is deposited in NoSQL databases, for review by medical staff when needed.

Unstructured data analytical capability: An analytical process in a big data management system starts by acquiring data provided from both inside and outside the healthcare sectors. After unstructured data has been gathered across multiple healthcare units, it is stored in a HDFS and NoSQL database that maintain it until it is called up in response to users' requests. The ability to analyze unstructured data plays a pivotal role in the success of big data in healthcare settings since 80% of health data is unstructured.

Analytical capability for patterns of care: Analytical capabilities in healthcare can be used to identify patterns of care and discover associations from massive healthcare records, thus providing a broader view for evidence-based clinical practice.

Decision support capability: Decision support capability emphasizes the ability to produce reports about daily healthcare services to aid managers' decisions. In general, this capability yields shareable information and knowledge such as historical reporting, executive summaries, drill-down queries and time series comparisons. Such information can be utilized to provide a comprehensive view to detect advanced warnings for disease surveillance and to develop personalized patient care.

Predictive capability: Predictive capability is the ability to apply diverse methods from statistical analysis, modeling, machine learning, and data mining to both structured and unstructured data to determine future outcomes. Predictive capabilities can reduce the degree of uncertainty and enable managers to support preventive care. The Texas Health Harris Methodist Hospital Alliance, for example, analyzes information from medical sensors to predict patients' movements and thus provide needed services more efficiently, Wang et al., 2015.

2.5 Big data based on graph databases

Despite Relational Database Management System (RDBMS) is the most popular and used in academic research, as well as industrial setup, graph databases regained interest among the researchers Kaliyar, 2015. Indeed, there has been developed a large number of systems for handling graph-like data like; social, biological, and other network.

Graph databases are the best for dealing with complex, semi-structure, and densely connected data. Generally, graph databases are useful when we are more interested in relationships between data than in the data itself, Angles, 2012. Graph database can traverse any number of relationships between entities, and they are efficient in retrieving relevant information after scouring several entities and relationship††. In the last time, there has been an increasing work on graph databases; from the current implementations, there are Neo4j, Allegro Graph, DEX, Hyper Graph DB, Infinite Graph and Sones.

According to Angles, 2012, a set of features is proposed in the literature in order to evaluate the data model provided by each graph database, that can be summarized as follows:

Graph data structures: the data structure of graph databases is defined around the notions of graphs, nodes and edges. There are four graph data structures; simple graphs, hyper graphs, nested graphs and attributed graphs, Angles, 2012. The basic structure is a simple flat graph defined as a set of nodes (or vertices) connected by edges. Nodes in graph may represent heterogeneous entities. A Hyper graph extends this notion by allowing an edge to relate an arbitrary set of nodes (called a hyper edge). A nested graph is a graph whose nodes can be themselves graphs (called hyper nodes). Attributed graphs are graphs where nodes and edges can contain attributes for describing their properties. Simple graph and attributed graph are the most supported by graph databases. Other features are considered which are directed or undirected edges, labeled or unlabelled nodes/edges, and attributed nodes/edges (i.e., edges between edges are possible). Graph data modeling can represent entities, properties and relations at both instance and schema levels.

Query languages: There is not proposal for a standard query language for graph databases. Some of the graph databases support predefined languages, such as Allegro Graph that supports SPARQL the standard query language for RDF; In contrary Neo4j is developing Cypher, a query language for property graphs.

Integrity constraints: Integrity constraints are general statements and rules that define the set of consistent database states, or changes of state, or both. However, integrity constraints are poorly studied in graph databases.

3. Motivation of graph databases in healthcare

According to Kaur and Rani, 2013, NoSQL databases enable programmers to model the data closer to the format used in their application domain. Graph databases which stores data as nodes and edges are the most appropriate NoSQL databases to model data in health care because these data contain as many relations between them as the amount of data themselves.

Healthcare systems based on graph databases provide a holistic and unified view of health data which helps doctors to diagnose and to predict diseases more quickly; consequently graph databases face an important big data challenge which is the data representation, Kaur and Rani, 2015; Ling et al., 2014.

Graph databases serve as multiple de-normalized tables which can avoid generating and replicating a large number of tables, thus reducing complexity in a database and enhancing data accessibility, Park et al., 2014. In addition, they have an intuitive query structure which facilitates the management, user validation, and

†† https://www.techopedia.com/2/31969/trends/big-data/graph-databases-a-new-way-of-thinking-about-data

exploration of the analytic intent of the query Park et al., 2014. Directly ,they can handle a wide range of queries that would otherwise require deep join operations in normalized relational tables, and performs well for some queries such as those supporting relationship mining, Park et al., 2014.Graph databases are also recommended to represent the temporal ordering of events while this type of queries is complex in relational database; for example a trajectory of patient's diseases, which is a set of records put together in chronological sequence, Sen et al., 2017

4. .Surveyed papers

Among the research work on big data, we have selected those exploiting graph databases in order to ensure above mentioned big data capabilities by storing, visualizing and analysing massive healthcare data. The first published works date from 2014, that's why the adoption of graph databases in healthcare sector is fairly recent. These works can be classified into three groups according to their objectives, namely, descriptive analytics, predictive analytics; preventive medicine and Large Scale Healthcare System. Descriptive analytics provides the ability to describe the data in summary from for exploratory insights and to answer "what has happened in the past?" questions. Predictive analytics allow users to predict or forecast the future for a specific variable, based on the estimation of probability, see Wang and Hajli, 2017. Preventive medicine is to take appropriate measures when identifying individuals having risk of developing a disease, see Khan et al., 2016 and Large Scale Healthcare System that refers to a system ensuring both of efficient data management and data services Park et al., 2014.

4.1. Descriptive and predictive analytics: GEMINI (Ling et al. 2014)

Objective: The objective of GEMINI is to respond to predicative tasks such identifying patients at high risk of developing heart disease in the near future, or predicting the probability that patients would re-admit into hospital within 30 days.

Data and structuring: GEMINI extracts clinical data from the CCDR[‡‡] of the national university hospital which has structured sources containing EMR. To interpret the unstructured data, GEMINI uses a well-known medical knowledge base UMLS[§§] and NLP engines.

Architecture of framework: The system consists of two components: PROFILING and ANALYTICS. The **PROFILING** component extracts data of each patient from various sources and stores them as information in a patient profile graph. The patient profile graph provides a holistic and unified view of a patient's clinical data. The **ANALYTICS** component analyses the patient profile graphs to infer implicit information and extract relevant features for the prediction tasks.

4.2. Descriptive analytics and preventive medicine

Framework to understand chronic disease progression (Khan, Uddin, et Srinivasan 2016)

Objective: The aim of the framework is to understand chronic disease progression in order to enable stakeholders making preventive measures.

Data and structuring: Hospitals, during the course of patient's admission and upon discharge, report the

[‡‡] Computerized Clinical Data Repository
[§§] Unified Medical Language System

detailed information in standard format to government departments and respective private health funds. In order to find the health trajectory of chronic disease patients and assess potential risk of developing disease for new patients, patients' admission records are analysed including length of stay; diagnose information, item-wise billing codes and patient's Diagnosis Related Group for each admission episode. The next step is to understand the semantics of the admission data and utilize them to develop a network that can represent the trajectory of chronic disease patients.

Architecture of framework: The framework is divided in 3 parts:

a. Part "a" collects and analyses patients' admission records and then understands the semantic of the admission data;

b. Part "b" creates baseline network to find patients' trajectory. This network represents typical sequence of diagnoses and comorbidity of chronic disease patients. The first phase of creating this network is Patient filtering which is essentially identifying chronic disease patients. And then statistical aggregation generates a graph of chronic disease patient's typical health trajectory based on admission history of chronic disease patients;

c. Part "c" finds the similarity between the Baseline Network of patients with chronic conditions and medical history of a new patient (not diagnosed with chronic disease). The method is named as Longitudinal Distance Matching, which uses sequential phases of rule-based, graph theory and social network analysis methods.

Portinari (Sen et al. 2017)

Objective: Portinari has as aim to explore and visualize future trajectories of patients who have undergone a specific sequence of screening exams in order to personalize cervical cancer screening and consequently reducing the number of cancer cases.

Data and structuring: Data is extracted from a socio-technical system for cervical cancer which is a system that identifies automatically individuals at risk of developing the disease and invites them for a screening exam; and available in the form of events in the life of patients stored in transaction records. In cervical cancer screening an event corresponds to attendance to exam, such as a Human Papilloma Virus, test along with a date and type of diagnosis.

Architecture of framework: The framework is divided in two components, namely:

Graph database of screening events: The events are transformed from transaction records into sequences of connected events for individual patients in a graph database implemented in Neo4J

Portinari: It is a web-based data exploration tool, to explore and visualize individual trajectories by querying the graph database. Portinari automatically generates future trajectories of patients who underwent the input sequence of exams and diagnosis by matching similar patients in the graph database. Portinari visualizes the outcome as a Sankey Diagram.

4.3. Large Scale Healthcare System: Framework for efficient data management and data services (Park et al. 2014)

Objective: Authors aimed to construct healthcare graph database; from a normalized relational database using the proposed "3NF Equivalent Graph" (3EG) transformation and to evaluate the performance of queries that require deep join operations over a relational database and its equivalent graph representation.

Data and structuring: Data are stored in 3NF relational healthcare database.

Architecture of framework: it consists of 3EG transform which is a graph database design rationale that constructs a graph database from an existing normalized database based on a group of conversion rules.

5. Analysis and discussion

Table 1 presents a comparison between healthcare's systems presented in section 4, based first on the objective of each work. Then, they are analysed according to the proposed architecture of big data; taking into consideration all the layers presented in section 2 except big data governance layer. At last, the systems were compared according to the structure of the graph.

Table 1. Comparison between healthcare systems based on graph database and big data technologies.

		(Ling et al. 2014)	(Khan, Uddin, et Srinivasan 2016)	(Sen et al. 2017)	(Park et al. 2014)
Objective		Predicative analytics	Prevention of chronic diseases	Prevention of cancer	Large scale healthcare systems
Data structure	Structured data	×	×	×	×
	Unstructured data	×			
Data sources	EMR	×	×	×	×
	Social networks				
	Captured data				
Data transformation	NLP	×			
Data storage	Graph data base management system			Neo4j	Neo4j
Data analysis	Hadoop Map Reduce	×			
	Stream processing				
	In-database analytics				
Information exploration layer	Visualization reporting	×	×	×	
	Real-time monitoring				
	Clinical decision support				
Graph Data Structure	Simple graph	×	×	×	×
	Hyper graph				
	Nested graph				
	Attributed graph	×	×	×	×
	Node Labelled				
	Node attribution		×	×	
	Edge Directed	×	×	×	
	Edge Labelled	×		×	
	Edges attribution		×	×	

By analysing Table 1, we draw the following conclusions:

- All the papers exploit structured data provided by EMR systems because it contains efficient and reliable data from which valuable information can be extracted.
- Ling et al., 2014 exploited unstructured data from doctor’s notes. Due to the complexity of this data, authors used several NLP techniques and UMLS dictionary to understand it and to construct the patient’s profile graph from this data and had to manage some limitations such as ambiguous mappings, missing mappings and relationships.
- Despite the existence of so much graph database management system, Neo4j is the most used because of many reasons. Neo4j is open source; it has an API and a query language (Cypher) so it is easy to handle and to query the database.
- Most of the works presented a visualization reporting as an output which is very helpful for doctors to draw conclusions about their patients and to predict future events. But the only visualisation was network graph for individuals or events when other dashboards may help a lot in statistic mostly for predictive tasks.
- None of the papers presented a real-time monitoring or clinical decision support which may be beneficial and helpful for descriptive analytics, predicative analytics and preventive medicine by sending notifications for patients having risk for developing a disease for example or determining the appropriate exam or medication.
- All of the papers opt for simple and attributed graph because those two types of graph are supported by most of graph databases management systems.
- Relationships in health data is as important as nodes, it contains relevant information that’s why many papers presented directed edges with attributes and labels.

6. Conclusion

The main contribution of this paper is to initiate a gap analysis related to graph databases and big data technologies in the healthcare area. In this sense, the first published works date from 2014, that’s why adoption of graph databases in the healthcare sector is fairly recent, despite the benefits it gives, the research is not completed.

The analysed papers raised some big data solutions in healthcare like descriptive and predictive analytics, preventive medicine and implementing large-scale system. However, there are other aspects, either in big data or healthcare area, are not yet tackled in the context of graph databases which may be the topic of future work, such as prescriptive analytics. This analysis requires the combination of optimization, machine learning, simulation and heuristics-based predictive modelling technique. The prescriptive analysis aims mainly to offer the optimal solutions or possible courses of action to help users understand what to do in the future while predicting risk of developing a disease, Wang and Hajli, 2017.

References

Angles, Renzo. 2012. « A Comparison of Current Graph Database Models ». In , 171‑77. IEEE. https://doi.org/10.1109/ICDEW.2012.31.

Asare-Frempong, Justice, et Manoj Jayabalan. 2017. « Exploring the Impact of Big Data in Healthcare and Techniques in Preserving Patients’ Privacy », 8 août 2017. http://paper.ijcsns.org/07_book/201708/20170819.pdf.

Cyganek, Boguslaw, Manuel Grana, Andrzej Kasprzak, Krzysztof Walkowiak, et Michal Wozniak. 2015. « Selected aspects of electronic health record analysis from the big data perspective ». In , 1391‑96. IEEE. https://doi.org/10.1109/BIBM.2015.7359881.

Ercan, Mehmet Zahid, et Michael Lane. 2014. « Evaluation of NoSQL databases for EHR systems », décembre 2014.

Fraczek, Konrad, et Malgorzata Plechawska-Wojcik. 2017. « Comparative Analysis of Relational and Non-relational Databases in the Context of Performance in Web Applications ». In Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation, édité par Stanisław Kozielski, Dariusz Mrozek, Paweł Kasprowski, Bożena

Małysiak-Mrozek, et Daniel Kostrzewa, 716:153‑64. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-58274-0_13.

Heinrich, Adrienne, Aizea Lojo, Ernestina menasalvas, et Wilfried Verachtert. 2016. « Big Data technologies in healthcare: Needs, opportunities and challenges ».

Kaliyar, Rohit kumar. 2015. « Graph databases: A survey ». In , 785‑90. IEEE. https://doi.org/10.1109/CCAA.2015.7148480.

Kaur, Karamjit, et Rinkle Rani. 2013. « Modeling and querying data in NoSQL databases ». In , 1‑7. IEEE. https://doi.org/10.1109/BigData.2013.6691765.

———. 2015. « Managing Data in Healthcare Information Systems: Many Models, One Solution ». Computer 48 (3): 52‑59. https://doi.org/10.1109/MC.2015.77.

Khan, Arif, Shahadat Uddin, et Uma Srinivasan. 2016. « Adapting Graph Theory and Social Network Measures on Healthcare Data: A New Framework to Understand Chronic Disease Progression ». In , 1‑7. ACM Press. https://doi.org/10.1145/2843043.2843380.

Krishnan, Shankar M. 2016. « Application of Analytics to Big Data in Healthcare ». In , 156‑57. IEEE. https://doi.org/10.1109/SBEC.2016.88.

Ling, Zheng Jye, Quoc Trung Tran, Ju Fan, Gerald C. H. Koh, Thi Nguyen, Chuen Seng Tan, James W. L. Yip, et Meihui Zhang. 2014. « GEMINI: An Integrative Healthcare Analytics System ». Proceedings of the VLDB Endowment 7 (13): 1766‑71. https://doi.org/10.14778/2733004.2733081.

Mathew, Prabha Susy, et Anitha S. Pillai. 2015. « Big Data solutions in Healthcare: Problems and perspectives ». In , 1‑6. IEEE. https://doi.org/10.1109/ICIIECS.2015.7193211.

Park, Yubin, Mallikarjun Shankar, Byung-Hoon Park, et Joydeep Ghosh. 2014. « Graph databases for large-scale healthcare systems: A framework for efficient data management and data services ». In , 12‑19. IEEE. https://doi.org/10.1109/ICDEW.2014.6818295.

Raghupathi, Wullianallur, et Viju Raghupathi. 2014. « Big Data Analytics in Healthcare: Promise and Potential ». Health Information Science and Systems 2 (1). https://doi.org/10.1186/2047-2501-2-3.

Sen, Sagar, Manoel Horta Ribeiro, Racquel C. De Melo Minardi, Wagner Meira, et Mari Nygard. 2017. « Portinari: A Data Exploration Tool to Personalize Cervical Cancer Screening ». In , 37‑46. IEEE. https://doi.org/10.1109/ICSE-SEIS.2017.6.

Wang, Yichuan, et Nick Hajli. 2017. « Exploring the Path to Big Data Analytics Success in Healthcare ». Journal of Business Research 70 (janvier): 287‑99. https://doi.org/10.1016/j.jbusres.2016.08.002.

Wang, Yichuan, Leeann Kung, Chaochi Ting, et Terry Anthony Byrd. 2015. « Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care ». In , 3044‑53. IEEE. https://doi.org/10.1109/HICSS.2015.368.

Yaqoob, Ibrar, Ibrahim Abaker Targio Hashem, Abdullah Gani, Salimah Mokhtar, Ejaz Ahmed, Nor Badrul Anuar, et Athanasios V. Vasilakos. 2016. « Big Data: From Beginning to Future ». International Journal of Information Management 36 (6): 1231‑47. https://doi.org/10.1016/j.ijinfomgt.2016.07.009.

Zillner, Sonja, et Sabrina Neururer. 2016. « Big Data in the Health Sector ». In New Horizons for a Data-Driven Economy, édité par José María Cavanillas, Edward Curry, et Wolfgang Wahlster, 179‑94. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-21569-3_10.

.