Volume XVI, Issue (1) (2020) 42-51

International May Conference on Strategic Management

GRAPH DATABASE MODELING OF A 360-DEGREE E-CUSTOMER VIEW IN B2C E-COMMERCE

Ilija Hristoski[1*], Tome Dimovski[2]

[1] *"St. Kliment Ohridski" University – Bitola, Faculty of Economics – Prilep, N. Macedonia;*
[2] *"St. Kliment Ohridski" University – Bitola, Faculty of ICTs – Bitola, N. Macedonia*

Abstract: For every B2C e-Commerce company, one of the major hurdles is the challenge of tracking the digital footprints of each e-Customer's activities during their online shopping sessions. As online competition becomes fiercer over time, online retailers face increasingly more sophisticated e-Customers. Knowing their buying habits and online shopping behaviors, which is a basic premise for building any strategies vis-à-vis retaining current and attracting new e-Customers, creates great opportunities for those who are capable of following and capturing relevant data about their e-Customers' digital trails. Usually part of contemporary CRM systems, the digital profile of an e-Customer, also known as 'a 360-degree e-Customer view', represents a collection of all e-Customers' data in one place. In this paper, a graph database modeling framework for constructing a 360-degree e-Customer view is proposed, with a single aim of exploring the possibilities of using NoSQL graph databases in storing highly relational data reflecting the complex interactions between e-Customers and a particular B2C e-Commerce website during online shopping sessions. The modeling framework is based on the utilization of a Customer Behavior Model Graph (CBMG) and is being implemented in Neo4j. The resulting graph database model represents a solid basis for answering a plethora of CRM-related questions.

Keywords: 360-degree view, e-Customers, graph databases, B2C e-Commerce, Customer Behavior Model Graph (CBMG), Neo4j

1. INTRODUCTION

The rapid growth and use of ICTs, particularly the broad proliferation of the Web 2.0 paradigm in the last two decades, have dramatically changed conventional ways of doing business throughout the world. The e-Commerce model has become a deciding factor both for the survival and prosperity of businesses in the contemporary global surrounding. Going online has provided businesses with plenty of new possibilities, but several possible obstacles to new market approaches have also been revealed. At the same time, the way how people shop has witnessed a seismic shift, too. The advent of the B2C e-Commerce paradigm has fundamentally changed the shopping behavior of millions of people worldwide since the modern ICTs have transformed every aspect of the sales process, including sourcing, browsing, searching, recommending, choosing, comparing, checking, ordering, receiving…

* *Corresponding author: ilija.hristoski@uklo.edu.mk*

This has contributed to a great deal of continual transformation of e-Customers, who became highly sophisticated. Though e-Commerce has provided more opportunities than ever before, it provides a very individual form of shopping for e-Customers. Each online shopper is different, displaying his/her patterns of behavior. Having minded this notion, it is of the utmost importance for online retailers to meet various e-Customers' expectations, and to induce positive online shopping experience and satisfaction. For online businesses, the best business strategy of all is, without question, the production of content e-Customers who might eventually produce even more e-Customers. Satisfied e-Customers are the only prerequisite to attract and then retain new e-Customers in the long run.

According to Roy H. Williams, a famous American businessman, "the first step in exceeding customer's expectations, is to know those expectations." In the current, so-called 'digital economy', making organizations more customer-centric becomes a crucial task, because the more they know about their customers, the less they need to worry about losing them. The more they know about their customers, the more they can provide to them the information that is increasingly useful, relevant, and persuasive. The so-called 'age of the customer', representing a transition in power from institutions and organizations towards customers, has changed and continues to change, the rules of business (Forrester.com, 2011). The first step in managing the value, experience, and expectations of customers successfully is to know who the customers are. Therefore, the design and construction of a 360-degree view of customers and their behavior must be prioritized, since "companies that make extensive use of customer analytics are more likely to have a considerable impact on corporate performance, outperforming its competitors" (Fiedler et al., 2016). In that context, a relevant, personalized, and ubiquitous customer service can be achieved solely by undertaking actions such as identification, knowing, evaluating, developing, customizing, retaining, and anticipating (Casariego Sarasquete, 2017, pp. 57–58).

The part '360-degree' denotes 'complete' or 'all-around', whilst the part 'view' refers to the ability to see something from a particular place or angle. Therefore, the term '360-degree view of a customer' suggests the ability to use the best available and most relevant information about each customer to enhance sales, marketing, and servicing decisions.

Recognizing the need for gathering all the e-Customers' interactions with a particular virtual store during their online shopping sessions, as well as the necessity to understand and learn from their online behavior, thus being able to anticipate their next moves, needs or expectations, the paper focuses on developing a framework for building a 360-degree view of e-Customers, based on the utilization of Customer Behavior Model Graphs (CBMGs) and NoSQL graph databases.

The rest of the paper is organized as follows. Section 2 provides an overview of some of the most prominent research, including white papers, made on the concept of a '360-degree view of customers' during recent years. The corresponding graph database model, based on the Customer Behavior Model Graph (CBMG) and the graph database model of an arbitrary e-Commerce website, is being proposed in Section 3, whilst Section 4 provides illustrative examples of how the graph database, implemented in Neo4j, can be effectively used for answering several CRM-related questions using Cypher Querying Language (CQL). The last section concludes.

2. RELATED RESEARCH

Due to the increasing significance, the research on the 360-degree view of customers is becoming a focusing point of a growing number of studies worldwide. What follows is just a glimpse of some of the most prominent research made lately on this hot topic.

Barker (2011) points out that the need to collect and manage data in a bid to recognize consumers at any touchpoint is greater than ever. By comparing internal and external customer data integration management, the article looks at the cutting-edge techniques for gathering and maintaining data both on- and off-line, while taking into account single customer databases and 360-degree views of customers. The author concludes that better data quality management can maximize value from customers' database records, and clearer and more accurate decisions can be made, which can ultimately improve profitability and revenues.

Recognizing the fact that there is a significant gap regarding the usage of Customer Knowledge Management (CKM) in practice, Vasireddy (2016) attempts to address the complex research problem of how to manage customer knowledge to achieve a 360-degree view of customers in organizations, by proposing a comprehensive CKM practice framework, based on the findings from the conducted empirical studies as well as from knowledge gained from the existent CKM literature overview.

Shahina et al. (2016) focus on the implementation of a customer-centric approach vis-à-vis e-Commerce-based friends' recommendation systems using the Neo4j graph database, thus enhancing their performances in terms of time and complexity.

Satish & Yusof (2017, p. 278) refer to the 360-degree view of customers from the perspective of Big Data analytics for enhanced customer experiences with crowdsourcing, as a means for showcase customer's interaction history with the corresponding outcomes. This can help to know the level of customer satisfaction and dissatisfaction with various products. From the crowdsourcing perspective, it should point out both positive and negative outcomes of product promotions, customer loyalty, and retention. The visualizations of the 360-degree view of customers should be able to give recommendations on how to promote products and new customer acquisitions to enhance customer support and co-learning of different aspects, as well as to troubleshoot errors and do anomaly detections, to conduct omnichannel pricing and promotions, and payment flow analysis.

According to Casariego Sarasquete (2017), "Being able to anticipate your customer's behavior is the holy grail for every business leader." The observed customer's behavior is made up of repetitive transactions and recurring purchases, as well as navigation and interactions through digital properties, channels, devices, applications, and social networks that can now be tracked and stored. The research establishes the foundation for a common data representation model of tracked customer behavior in a form of a 360-degree holistic view of the customer and its behavior that can be used effectively by organizations to represent any given customer-centric business model, thus allowing the analysis of the most popular marketing problems like segmentation, cross-selling, retention, etc.

Finsterwalder (2018) investigates the 360-degree view of actor engagement in service co-creation, pointing out the need for interpreting it as an encompassing concept with 'receptive' (i.e. psychological) as well as 'transmissive' (i.e. behavioral) properties.

3. BUILDING A MODEL OF A 360-DEGREE VIEW OF AN E-CUSTOMER

The majority of e-Customers that visit a particular e-Commerce website can be classified into two, three, or more classes according to their similar behavior patterns they exhibit during online shopping sessions. To describe the online shopping behavior of specific classes of e-Customers, special types of graphs, known as Customer Behavior Model Graphs (CBMGs) have been introduced (Menascé & Almeida, 2000; Menascé & Almeida, 2000). CBMGs are graph-based models that characterize web shopping sessions of e-Customers belonging to a specific class, while they are visiting a particular virtual store. Put differently, they capture the navigational patterns of e-Customers through a particular e-Commerce website, as viewed from the web server perspective (Figure 1).

Each particular e-Customer exhibits various behaviors during various online shopping sessions, i.e. no two online shopping sessions, even of the same e-Customer, are identical. The specific online shopping behavior of an e-Customer determines both the dynamics (i.e. the frequency and the randomness) and the structure (i.e. the order) of the invoked e-Commerce functions, such as LOGIN, REGISTER, LOGOUT, SEARCH, BROWSE, BUY_NOW, etc. It should be notified that, due to simplicity reasons, the simplified CBMG shown in Figure 1 does not depict many well-known e-Commerce functions that can be found with contemporary online stores, such as SAVE_FOR_LATER, MAKE_OFFER, ADD_TO_WATCHLIST, etc.

For instance, having minded a CBMG of a hypothetical generic e-Commerce website (Figure 1), a specific e-Customer who knows what to buy and is highly determined to make an online purchase, may well invoke a specific set of e-Commerce functions in the following order:

ENTER_PAGE → LOGIN → SEARCH → VIEW_ITEM → BUY_NOW → LOGOUT → EXIT_PAGE

On the other hand, previously, the same e-Customer, without knowing what to buy, and, therefore, without being determined to make an online purchase, could invoke e-Commerce functions in the following sequence:

ENTER_PAGE → BROWSE → VIEW_ITEM → … → BROWSE → VIEW_ITEM → SEARCH → VIEW_ITEM → BROWSE → VIEW_ITEM → … → BROWSE → VIEW_ITEM → EXIT_PAGE

Finally, the same e-Customer, who knows what to buy, but is rather reluctant, because he/she is not determined to make an online purchase, could invoke a specific set of e-Commerce functions in the following order:

ENTER_PAGE → SEARCH → VIEW_ITEM → LOGIN → ADD_TO_CART → VIEW_CART → REMOVE_FROM_CART → EXIT_PAGE

Each particular CBMG, specific to an e-Commerce website, is comprised of N states, where state #1 is always the ENTRY state (designated by letter 'E' in Figure 1), and state #N is always the EXIT / LOGOUT state (designated by letter 'X' in Figure 1), whilst the states 2, 3, …, N−1 correspond to the states HOME (1), BROWSE (2), …, REMOVE_FROM_CART (11), respectively. Besides the characteristic set of states, a CBMG is being also described by a set of possible transitions between two particular states i and j, designated by directed arcs from state i to state j. The set of states and the set of possible transitions refer to the static aspect of a CBMG since they reflect the structure of the e-Commerce website and does not depend on the way e-Customers access and use it. All e-Customers of a particular e-Commerce website share the same static aspect of the CBMG.

The $N \times N$ transitional probability matrix $P = p[i, j] = p_{i \cdot j}$, whose elements are the probabilities of transiting from state *i* to state *j* in one step, represents the dynamic aspect of a CBMG. In Figure 1, such probabilities, which denote, in fact, relative frequencies of invoking specific e-Commerce functions, are being designated in a form of labels $p_{i \cdot j}$, assigned to each directed arc between certain pairs of states in the CBMG.

The proposed graph database model for gaining a 360-degree e-Customer view of a particular online retailer website is based solely on the static aspect of its specific CBMG. The e-Customer-centric graph database model, resembling the interaction between a particular e-Customer and a hypothetical generic B2C e-Commerce website, is depicted in Figure 2. All N states of a CBMG shown in Figure 1 are being transformed into relationships in the graph database model, starting from a single central node, labeled as INTERNET_USER (Figure 2), and ending into several other nodes' types, representing corresponding entities with specific attributes, which are constitutive parts of the generic e-Commerce relational data model proposed by Williams (2009). This way, the invocations of various e-Commerce functions, which reflect the interaction between a specific e-Customer and the online retail store during online shopping sessions, are naturally represented as relationships in the graph database model. The semantic power and the semantic context of the proposed graph database model rely entirely on the relationships' names, directions, and attributes since graph databases are purpose-built to store and navigate relationships. Relationships are 'first-class citizens' in graph databases, and most of the value of graph databases is derived from these relationships. In this particular case, the relationships connecting pairs of nodes are more valuable than the data itself because they represent the invocations of e-Commerce functions, i.e. they reflect the e-Customer's interaction with the e-Commerce website.

The resulting graph database could address "one of the great macroscopic business trends of today: leveraging complex and dynamic relationships in highly connected data to generate insight and competitive advantage", because "the ability to understand and analyze vast graphs of highly connected data will be key in determining which companies outperform their competitors over the coming decade" (Robinson et al., 2015).

All the relationships represented in Figure 2, besides their specific attributes, include a set of three custom attributes: date_time (a timestamp), session_GUID (a globally unique identifier of each particular online shopping session), and IP_address (the IP address of the Internet user/e-Customer interacting with the e-Commerce website). These attributes are the key ones in the process of traversing the graph during the execution of various queries that contribute to the completion of 360-degree views of e-Customers. The date_time attribute stores the date and time information in a format 'yyyy-mm-dd hh:mm:ss.sss' according to the ISO 8601 international standard. The session_GUID attribute stores a 128-bit integer number, which is both big and distinctive enough to identify each online session uniquely. Each online shopping session starts at the moment when the Internet user accesses the e-Commerce webpage for the first time and lasts till the moment when he/she exits it or closes the window rendering the HTML code of the e-Commerce webpage. The IP_address attribute stores the IPv4 address of the Internet user who accesses and interacts with a particular e-Commerce store in dotted decimal notation.

As an example, Figure 3 displays an excerpt from the test graph database, implemented in Neo4j, a highly scalable and schema-free (NoSQL) graph database management system, offering ACID-compliant transactions and native graph storage and processing. The test graph database itself is built entirely using the graph database model presented in Figure 2 as a blueprint. It shows only the interactions of a particular e-Customer with a specific e-Commerce website during accomplishing three of his/her online shopping sessions, each comprised of a set of e-Commerce functions invoked in a specific order, being mentioned previously in this Section. For demonstration purposes, the e-Customer accesses the e-Commerce webpage from two distinct IP addresses, whilst all the values of nodes' and relationships' attributes have been chosen arbitrarily.

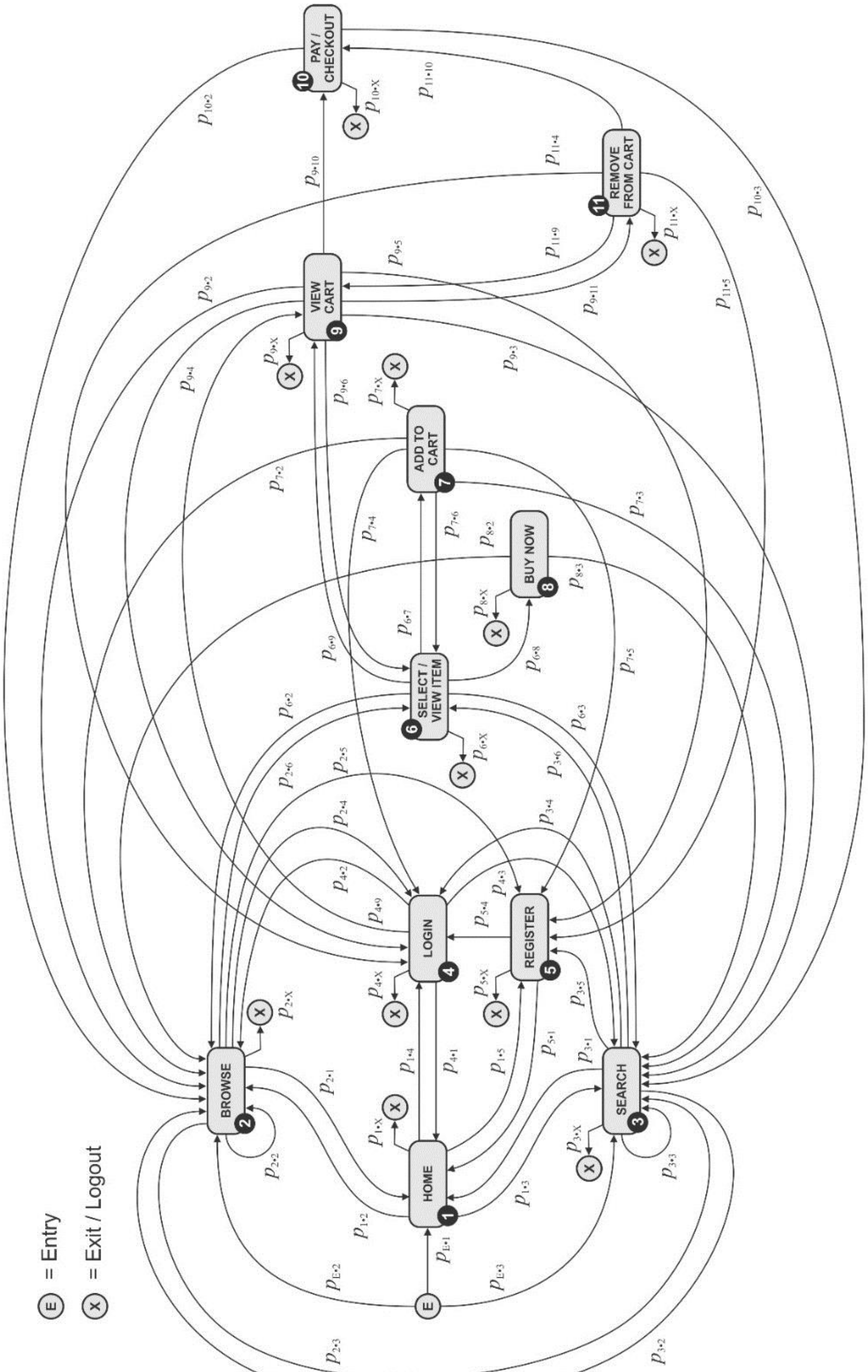

Figure 1. CBMG of a hypothetical generic B2C e-Commerce website (rotated)

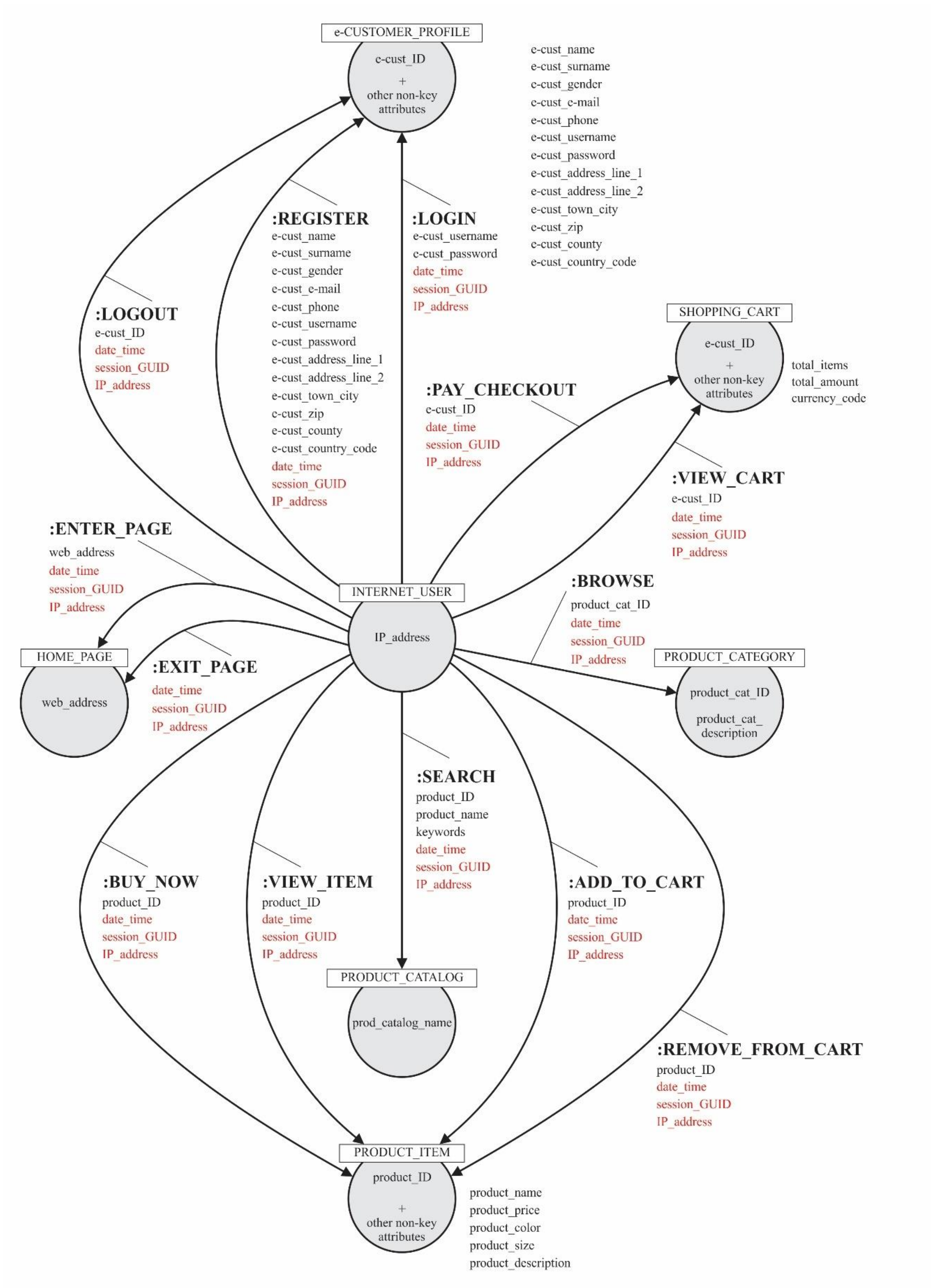

Figure 2. E-Customer-centric graph database model, resembling the interaction between e-Customers and a hypothetical generic B2C e-Commerce website

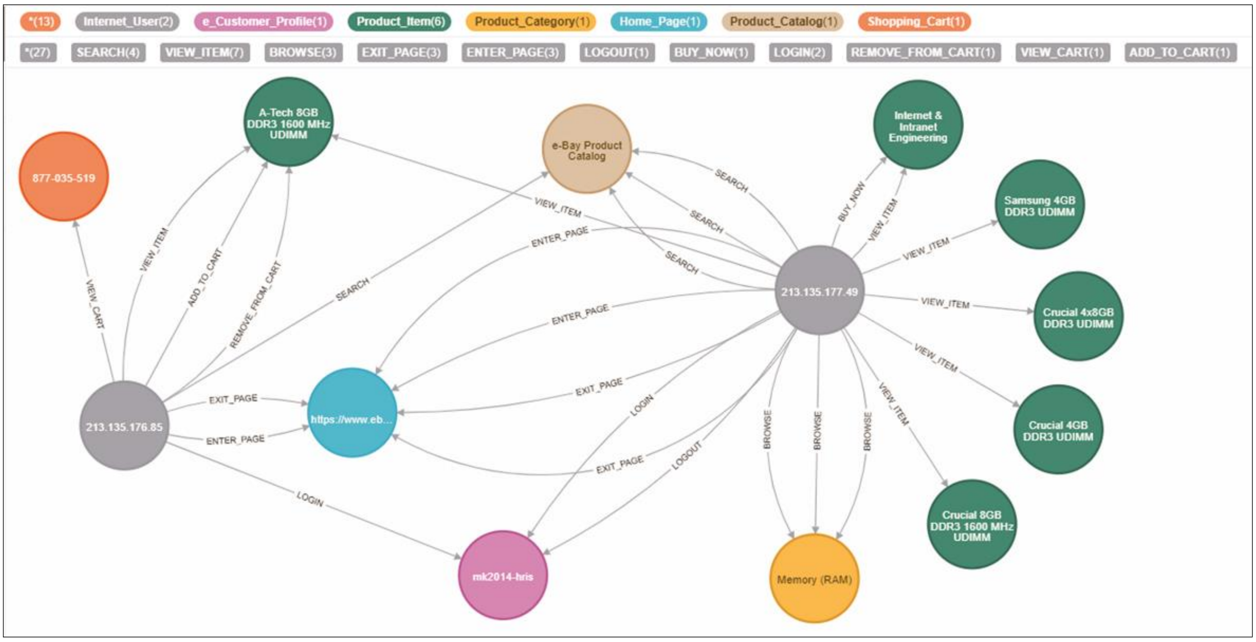

Figure 3. Excerpt from the Neo4j-based graph database, depicting the interactions between a specific e-Customer and a hypothetical B2C e-Commerce website during three online shopping sessions from two distinctive IP addresses

4. ANSWERING CRM-RELATED QUESTIONS USING A GRAPH DATABASE

Table 1 contains some of the most important CRM-related questions (first column), as well as the Cypher Query Language (CQL) programming code, for defining corresponding queries (second column) to address the posed questions vis-à-vis the implementation of the previously proposed graph database model in Neo4j.

Table 1. Examples of CRM-related questions and the CQL code needed for their addressing

CRM-related question	Cypher Query Language (CQL) code
How many times has e-Customer 'MK2014-hris' logged in to the e-Commerce webpage during August 2020?	`match (a:Internet_User)-[r:LOGIN]->(b:e_Customer_Profile) where r.e_cust_username="mk2014-hris" and left(r.date_time, 4)="2020" and substring(r.date_time, 5, 2)="08" return count(*) as number_of_logins`
What distinct IP addresses have e-Customer 'MK2014-hris' used to log in to the e-Commerce webpage so far?	`match (a)-[r:LOGIN]->(b) where b.e_cust_username="mk2014-hris" return distinct a.IP_address`
Knowing that e-Customer 'MK2014-hris' has logged in to the e-Commerce website from two IP addresses, 213.135.176.85 and 213.135.177.49, *what unique product categories has he/she browsed* so far?	`match (a)-[:BROWSE]->(b) where a.IP_address="213.135.176.85" or a.IP_address="213.135.177.49" return distinct b.product_cat_description`
Knowing that e-Customer 'MK2014-hris' has accessed the e-Commerce website from the IP addresses 213.135.176.85 and 213.135.177.49, *how many online shopping sessions has he/she accomplished* during August 2020?	`match (a:Internet_User)-[r]->(b) where (a.IP_address="213.135.176.85" or a.IP_address="213.135.177.49") and (left(r.date_time, 4)="2020" and substring(r.date_time, 5, 2)="08") return count(distinct r.session_GUID) as shopping_sessions`

Knowing that e-Customer 'MK2014-hris' has accessed the e-Commerce website from the IP addresses 213.135.176.85 and 213.135.177.49, *how many different products has he/she seen* during August 2020?	`match (a:Internet_User)-[r:VIEW_ITEM]->(b:Product_Item) where (a.IP_address="213.135.176.85" or a.IP_address="213.135.177.49") and (left(r.date_time, 4)="2020" and substring(r.date_time, 5, 2)="08") return count(distinct b.productID) as viewed_product_items`
How many items did e-Customer 'MK2014-hris' buy instantly so far during all online shopping sessions made from the IP address 213.135.177.49?	`match (a:Internet_User)-[r:BUY_NOW]->(b:Product_Item) where a.IP_address="213.135.177.49" return count(b.productID) as instantly_bought_product_items`
Knowing that e-Customer 'MK2014-hris' has accessed the e-Commerce website from the IP addresses 213.135.176.85 and 213.135.177.49, *what is the total amount he/she spent on buying products instantly so far*?	`match (a:Internet_User)-[r:BUY_NOW]->(b:Product_Item) where a.IP_address="213.135.176.85" or a.IP_address="213.135.177.49" return sum(b.product_price) as total_amount_spent_instantly`
Regarding all online shopping sessions of e-Customer 'MK2014-hris' made from the IP address 213.135.176.85, *what keywords he/she used while searching for products*?	`match (a:Internet_User)-[r:SEARCH]->(b:Product_Catalog) where a.IP_address="213.135.176.85" and r.keywords<>"" return r.keywords`
What was the online shopping session duration (in seconds) of e-Customer 'MK2014-hris', with a session_GUID = "4cc7de0930bf4e27b40bbc968f1e2fb7", made from the IP address 213.135.177.49?	`match (a:Internet_User)-[t:ENTER_PAGE]->(c:Home_Page) match (a:Internet_User)-[x:EXIT_PAGE]->(c:Home_Page) where a.IP_address="213.135.177.49" and t.session_GUID="4cc7de0930bf4e27b40bbc968f1e2fb7" and x.session_GUID="4cc7de0930bf4e27b40bbc968f1e2fb7" return duration.inSeconds(localtime(right(t.date_time,12)), localtime(right(x.date_time,12)))`

Numerous other CRM-related questions can be answered by a slight modification of the relatively simple CQL queries presented in Table 1. Moreover, answers to more complex questions (e.g. finding the exact order of invoked e-Commerce functions during a specific online shopping session) can be found using graph traversing algorithms like Breadth-First Search (BFS) or Depth-First Search (DFS), which are often a required first step for many other types of analyses.

5. CONCLUSION

The proposed framework for attaining a 360-degree view of an e-Customer is based on the transformation of states found within a CBMG of a specific e-Commerce website into relationships of a graph database model. In a NoSQL graph database implementation, each particular invocation of an e-Commerce function generates a relationship with its specific attributes that contain the complete information of what, when, and how. This way, the graph database can keep a digital trail of all e-Customers' actions during online shopping sessions, which is in line with the definition of the 360-degree view, according to which it is a complete full-circle view of who they are, and, importantly, it sheds light on every angle of their interaction with the virtual store. It is the biggest benefit of this approach. On the other hand, depending on the e-Customer's behavior, even a moderate online shopping session in terms of the duration and intensity of invoking e-Commerce functions can generate a huge number of relationships in a graph database. Repetitive online sessions accomplished by a single e-Customer can contribute to a dramatic increase in the amount of data and complexity of relationships stored within the graph database, which is the biggest disadvantage. However, this issue paves the way for engaging other contemporary technologies, such as Big Data analytics, having minded the fact that NoSQL databases, including graph databases, are constituent, yet fundamental parts of such technologies. The proposed approach offers huge potentials for answering practically all relevant questions regarding e-Customers and therefore it can be utilized as a solid basis for building powerful CRM information retrieval systems.

REFERENCES

Barker, D. (2011). Customer data integration: Reaching more consumers with certainty. Journal of Database Marketing & Customer Strategy Management, 18(3), 214–219.

Casariego Sarasquete, N. M. (2017). A Common data representation model for customer behavior tracking. Icono 14, 15(2), 55−91.

Fiedler, L., Großmaß, T., Roth, M., & Vetvik, O. J. (2016). Why customer analytics matter. Retrieved from http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/why-customer-analytics-matter (accessed August 16, 2020)

Finsterwalder, J. (2018). A 360-degree view of actor engagement in service co-creation. Journal of Retailing and Consumer Services, 40, 276–278.

Forrester.com (2011). Age of the Customer. Retrieved from https://go.forrester.com/age-of-the-customer/ (accessed August 16, 2020)

Menascé, D. A., & Almeida, V. A. F. (2000). Scaling for E-Business: Technologies, Models, Performance, and Capacity Planning, Upper Saddle River, NJ: Prentice-Hall PTR.

Menascé, D. A., & Almeida, V. A. F. (2002). Capacity Planning for Web Services: Metrics, Models, and Methods, Upper Saddle River, NJ: Prentice-Hall PTR.

Robinson, I., Webber, J., & Eifrem, E. (2015). Graph Databases: New Opportunities for Connected Data, Second Edition. USA: O'Reilly.

Satish, L., & Yusof, N. (2017). A Review: Big Data Analytics for enhanced Customer Experiences with Crowd Sourcing. Procedia Computer Science, 116 (2017), 274–283.

Shahina, C. P., Bindu, P. S., & Mariam Varghese, S. (2016). Enhancing the Performance of e-Commerce Solutions by Friends Recommendation System and Neo4j Database. International Journal on Cybernetics & Informatics, 5(2), 165−171.

Vasireddy, P. (2016). Customer Knowledge Management in organisations: Developing a practice framework to achieve360o view of customer. Ph.D. Thesis. University of Bolton, Bolton, United Kingdom.

Williams, B. (2009). A Data Model for e-Commerce. Retrieved from http://www.databaseanswers.org/data_models/e_commerce/ index.htm (accessed August 10, 2020)