



Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Modeling user behavior data in systems of engagement

Oliver Bent^a, Prasenjit Dey^b, Komminist Weldemariam^{a,*}, Mukesh K. Mohania^b^a IBM Research, Nairobi, Kenya^b IBM Research, Bangalore, India

HIGHLIGHTS

- We identified various issues in modeling systems of engagement personalized education context.
- We proposed a system comprising stream of data capturing, management, analytics and visualization.
- We leverage various analytic techniques (e.g. MapReduce) on Cloud to model systems of engagement.
- We evaluate the five characteristics of Big Data based on a pilot program in school district in Africa.

ARTICLE INFO

Article history:

Received 1 May 2015

Received in revised form

3 April 2016

Accepted 25 May 2016

Available online xxxxx

Keywords:

User engagement

User modeling

Instrumentation

Mobile sensors

ABSTRACT

The proliferation of mobile devices has changed the way digital information is consumed and its efficacy measured. These personal devices know a lot about user behavior from embedded sensors along with monitoring the daily activities users perform through various applications on these devices. This data can be used to get a deep understanding of the context of the users and provide personalized services to them. However, there are a lot of challenges in capturing, modeling, storing, and processing such data from these systems of engagement, both in terms of achieving the right balance of redundancy in the captured and stored data, along with ensuring the usefulness of the data for analysis. There are additional challenges in balancing how much of the captured data should be processed through client or server applications. In this article, we present the modeling of user behavior in the context of personalized education which has generated a lot of recent interest. More specifically, we present an architecture and the issues of modeling student behavior data, captured from different activities the student performs during the process of learning. The user behavior data is modeled and sent to the cloud-enabled backend where detailed analytics are performed to understand different aspects of a student, such as engagement, difficulties, and preferences and to also analyze the quality of the data.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The “Big Data” phenomenon has led to an awareness of ever increasing volumes of data of various forms be they structured, semi-structured or unstructured. Rapid and timely analytics on Big Data offers significant advantages to business and scientific communities. A recent study [1] predicts that in the future, organizations will have to deal with more and more unstructured and semi-structured data generated from sensors and devices. The same study also predicts that a majority of the workload in

organizations will comprise real-time analytics, using this data which is constantly being generated.

Similarly, enterprises are fully realizing the value of data and have begun to store it in Customer Relationship Management (CRM) systems, transactional systems, operational data stores and data warehouse systems. Such systems are designed as “System of Records” (SORs) and are used to derive business insights by analyzing the data while it is at rest in SORs. However, in the last few years there is a paradigm shift in building the next generation enterprise management systems from the traditional “Systems of Records” to the “Systems of Engagement” (SOE)—see, e.g. in [2–4]. In the industry of Education, the term SOE refers to the ability of systems through which students, teachers, employees, etc. directly interact or engage with peers or with larger backend systems. Systems of engagement may know a lot about user behavior and interaction as they interact or engage

* Corresponding author.

E-mail addresses: oliverbent@ke.ibm.com (O. Bent), prasenjit.dey@in.ibm.com (P. Dey), k.weldemariam@ke.ibm.com (K. Weldemariam), mkmukesh@in.ibm.com (M.K. Mohania).

<http://dx.doi.org/10.1016/j.future.2016.05.038>

0167-739X/© 2016 Elsevier B.V. All rights reserved.

through various channels such as learning apps and social media. This is due to several factors including new ways to interact with the systems through mobile devices, sensors storing contextual information, and disseminating data-as-a-service on cloud, etc. Thus, the powerful combination of mobile or tablet devices and sensors as the frontend for sensing user behavior and cloud as the backend for analysis across different application domains has led to the design of new architectures.

Personalized education [5] is one such application area which has seen the proliferation of mobile devices due to government initiatives, affordability, a large ecosystem of applications, and ease of use. These devices may in turn be used to understand student behavior, when they use the devices for learning, which can help teachers track their progress, understand their difficulties, and personalize their learning plans [6,7].

Student behavior data possesses all the generally cited key characteristics of “Big Data” including velocity, variety, veracity, volume, and value. Thus, in this paper, we first describe the characteristics of the user interaction, affective and cognitive data, and map them to Big Data characteristics [8]. We further outline how to model such data collected across different devices and applications and then how to analyze these vast data streams to derive actionable insights in the context of systems of engagements. We also outline some issues that are associated in building such systems that enable collecting and modeling fast moving data for efficient processing and analysis. In our system, we store the user interaction data as JSON objects in IBM Cloudant DBaaS (Database-As-A-Service) and process it using MapReduce functions for data analysis. Finally, using the dataset we generated in a limited pilot, we show experimental analysis of user behavior and engagement modeling for a single cohort of students and their teachers.

The major contributions of this article are as follows:

- We identify various issues in modeling systems of engagement in the context of education using the 4Vs (volume, velocity, variety and veracity) along with the additional term Value (Section 2).
- We propose an end-to-end system architecture that constitutes client side applications design, suited to capture the stream of end-user data, back-end data management and analytics, and visualization (Section 3).
- We leverage various analytics techniques (e.g. MapReduce) on cloud to model systems of engagement (Section 3).
- We evaluate the five characteristics of Big Data using a real-world dataset (Section 4).

2. Background and motivation

In this section, first we discuss the 4Vs (Volume, Velocity, Variety and Veracity) along with the additional term Value and outline various issues in modeling such data. We then highlight some of the privacy and security concerns around such systems of engagement data.

2.1. Characteristics of big data

User behavior data displays all the core 4V characteristics of Big Data and one additional V for Value.

Volume: When users consume information on a mobile and tablet devices, the various sensors embedded on these devices can be instrumented to collect rich and fine-tuned data. The activity of the users on the application also presents a lot of data which may not be of such great volume depending on sampling rates. The sampling rates for sensors such as accelerometer, camera, microphone, and temperature are very high and hence a high

volume of data is generated. Using such high volumes of generated data, the key research questions to understand are as follows: What data to process at the client and what to send to the cloud (server)? What level of detail of the data should we model? This will provide us with an understanding of what volume of data we need to store in the backend and the minimum data we can process on the device (client) for immediate decision making or what can be discarded. This falls in new kinds of distributed storage and analysis mechanisms such as in Cloudant—a distributed NOSQL JSON store. The large volume of the data requires distributed storage and analysis frameworks, such as MapReduce to extract insights.

Velocity: The stream of activity events, whether from sensors or the application on the mobile device, has high velocity. This has impact on storing the data. Some crucial decisions need to be made such as what samples can be dropped without affecting the overall value of the data? How can the cloud storage backend handle such rapid spurts of data? We need to devise buffering mechanisms at the device to make sure the cloud backend is able to handle the flow of data. Moreover, when backend network connectivity is an issue – a typical problem in resource constrained regions such as Africa and India – the client-side event management and buffering mechanisms should be robust enough to efficiently summarize and buffer this data and sync-up with the cloud when the network becomes available. Cloudant allows for such robust buffering and synchronization of high velocity data, with the Cloudant Sync library.

Variety: The variety of data in user behavior is very high. This aspect of the behavior data is the key to our decision to go for a NOSQL data store. We model each session of interaction of a user as a JSON document. However, as new devices with new sensors emerge, and new applications are required, the events that are captured will keep on changing and evolving. The types of data vary in their parameterization. For example, the camera data may contain bounding box coordinates, or the microphone may have dB level number, or the Play/Pause event may have just Yes/No (all with a timestamp). A traditional relational database schema would have been too constraining to model a priori all the possible events. The flexibility of a NOSQL document store allows for defining new events in the future while still being able to analyze the new as well as the old data for various events in different contexts. However, some interesting research questions emerge: How detailed should the model be for each event key? How to efficiently capture temporal events? How to model the unstructured data in a relational world? Should the entities be extracted and stored in relational schemas or should they be stored in a native form? The key here is the flexibility of capturing and modeling the data with continuous analysis to provide insights.

Veracity: The data from sensors are inherently unreliable. Sensors may not be of the same accuracy and sensitivity in all devices or may not be present at all. Still moreover as the user has control over what data is collected they may not allow data collection from embedded sensors at all. The data model that we create should be able to handle such aspects. NOSQL JSON document stores can create documents that have only those events in them and not the ones that are absent. This is a big advantage, since in relational databases this would have resulted in sparsely populated columns whenever some sensor data is corrupt or not available. In this behavior-data world, complete quality of all the data cannot be ascertained but the efficient storing, analysis, and modeling of the majority of the data is of importance.

Value: Another very important aspect of big data is the value of the data. Unless the data has value for the business problem one is trying to solve it is a futile exercise. Every aspect of collecting, transporting, and storing such large amounts of data continuously requires investment in infrastructure and people.

Download English Version:

<https://daneshyari.com/en/article/4950518>

Download Persian Version:

<https://daneshyari.com/article/4950518>

[Daneshyari.com](https://daneshyari.com)