



Journal of Knowledge Information Technology and Systems

ISSN 1975-7700 (Print), ISSN 2734-0570 (Online)

<http://www.kkits.or.kr>

A Study on Big Data Platform Architecture-based Conceptual Measurement Model Using Comparative Analysis for Social Commerce

Sung-Tae Hwang*

Department of IT Engineering, Hansung University

A B S T R A C T

With the Big Data growing in the popularity, more and more architectures are being proposed, but no unified solution is at hand. The main objective of this thesis is to specify and prepare a conceptual measurement model on which selected Big Data Platform Architecture will be mapped and compared. The conceptual measurement model is split into multiple components with the specification of responsibilities, and the their scope of the conceptual measurement model is set. All architectures mapped in a single structure are then evaluated and compared. Our research contains a brief overview of the Big Data concept and a summary of technologies used in evaluated architectures, if available. In addition, social commerce is a new paradigm infrastructure model of commercial market, which use of Web 3.0 technologies and social data to support their exchange activities. While it is popular, being a subset of commercial market, has been increasing tremendously since its introduction in 2012, there exists a general particular of research on their framework and applications' effectiveness, especially in areas beyond the general social commerce practices. This research develops a comprehensive social commerce framework in big data environment that has four key components. Then our research applies the Big Data Platform (BDP) to analyze the usability and adaptability with the framework by applying it to social commerce successful companies. Accordingly, we provide a set of metrics for social commerce-based Big Data Platform design. The proposed our framework and they are used to lead the design and evaluation in social commerce environment. This research makes a contribution to the social commerce in proposing a new design framework and also, we have to examine in providing inside for effective approach of concerning with their social business.

© 2020 KKITS All rights reserved

KEYWORDS : Social commerce, Commercial market, Web 3.0 technology, Social data, Big data platform, Social business

ARTICLE INFO: Received 10 September 2020, Revised 21 September 2020, Accepted 13 October 2020.

*Corresponding author is with the Department of IT Engineering, Hansung University, 116

Samseongyuoro-16gil Seongbuk-gu Seoul, 02876, KOREA.
E-mail address: sthwang@hansung.ac.kr

1. Introduction

Social Commerce is caused by the evolution of commercial market and the widespread use of social network-based data. In a various of systems and smart platform provide to commercial services with product data and customer historical record by online commerce market such as e-commerce, m-commerce or u-commerce [1]. Their commerce market is based on Amazon, eBay, Facebook, or Coupang and so on, these system is supported on recommendation obtained using the web community and social data. Nevertheless they were fast growth and increasing with social commerce what they want service, they did not satisfy about their commercial policy how to effectively manage to users for social service [2]. In other words, despite they made up design with online market place, they have a question what components need in a social commerce framework, what they want communications for personalized service, how they connect social channel with system to system, system to user or system to platform [3]. These are based on the integration of previous system to newly system and also, the existing knowledge of big data environment. Accordingly, our main idea propose a new framework and a set of fundamental principle for social commerce [4]. Moreover, we provide Big Data Platform architecture analysis for social commerce including four key elements [5-9].

In order to main contributions of our work, this paper are that we propose a users and service provider using BDP Architecture.

This paper is organized as follows. In Section 2 details the Big Data. Then Big Data Platform architecture is presented and discussed in Section 3. Finally, Section 4 concludes the paper by briefly summarizing the main points and proposing future work.

2. Big Data

There is no single formulation of what Big Data actually is. Generally, it is described as the data that enhances the processing capacity of traditional systems and cannot be processed or analyzed using conventional tools or processes. More and more organizations are receiving and collecting large amounts of data, but they often do not have the means of how to store and analyze the information. This spark of information overflow comes from multiple facts. Sensors are becoming a natural part of every device, and many sensors are capturing information on a constant frequency [10]. It is a trend nowadays to perform most of the activities on the Internet using different services for different needs, and people tend to generate massive volumes of data on social networks. However, the amount of information is only one of the aspects of Big Data. The data, especially from sensors, usually come in a raw, semistructured, or unstructured format. It is necessary to prepare the received data before it can be stored and analyzed. The analysis became more complex as well, requiring statistical, data mining, and machine learning to manage the Big Data.

2.1 5V Model

Three main characteristics of Big Data are defined and were mainly mentioned at the beginning. These characteristics are variety, velocity, and volume [11]. Later on, this list was extended with two more characteristics: Value and Veracity. Together, these characteristics are often referred to as 5V. The Volume indicates that the Big Data system can contain petabytes of records, and receive terabytes of data every hour. This number is expected to increase to zettabytes in the future. The Velocity refers to the speed of the information retrieval and the speed of the data flow in the system. The incoming data can be in different formats. Either structured, unstructured, semistructured, or a mix of them, the Big Data system should be able to facilitate them all, and that is the characteristic of Variability.

2.2 Hadoop

Hadoop is one of the most popular implementation frameworks developed by the Apache as an open-source project. It is in the Apache Software Foundation, and the code-base is written in Java. Hadoop is built on top of a distributed file system and is designed to scale up naturally with the number of available hardware. The inspiration came from Google's work on the GFS(Google File System) and the MapReduce paradigm. The main methodology is based on the function-to-data model rather than data-to-function model [12]. The biggest difference is in the

process, where the analysis programs are sent to their data and not the other way.

2.3 HDFS

Distributed file systems are file systems that deal with the storage on a network of machines. This functionality is necessary when the dataset cannot fit on the single physical machine and must be distributed across multiple machines. The management of data consistency and redundancy is even a bigger problem in Big Data environments [13]. Data in Hadoop are divided into smaller blocks and shared throughout the cluster. It is then easier to apply MapReduce functionality on smaller datasets, thus providing efficient processing in Big Data systems.

2.4 MapReduce

MapReduce is a programming paradigm concentrating on the clustered parallel data processing part of the Big Data system. It consists of two separate phases, firstly map phase takes a set of data on the input and converts it into multiple tuples, which are key-value pairs. For example, it can extract and transform some information from the raw data on the input, if only some of the values are useful for the second phase, the reduce. The reduce phase always comes after the map phase, and it has the output of the map job as the input. The reduce combines tuples on the input into smaller sets of tuples [14-16].

3. Big Data Platform Architecture

Big Data Platform architecture is a regular software architecture that concentrates on the Big Data domain. Since the Big Data is such a complex environment, it is more than necessary to specify an architecture before actually building the system. As mentioned before, the architecture should be evaluated as soon as possible, to prevent future issues. Currently, there are many different Big Data architectures specified, but no general solution is agreed as a standard. The lack of the standard in Big Data may be caused by the fact that it is a relatively new concept and there was simply no time to establish it, or by the complexity of the domain. Thus, it is complex to compare them in order to design a new architecture or to enhance existing ones. The enhancement can be done by comparing the existing architecture against some unified format and then evaluating the parts that were not found in the scope of the current architecture. During the evaluation, it has to be decided whether the part of the system that is currently not implemented can bring some additional value from the business or technical perspective. If yes, a guideline on how this component should work and cooperate with the rest of the system is needed. Moreover, technology has to be selected for the new part of the system that can be easily integrated with the existing solution.

4. Conceptual Measurement Model for BDP

The conceptual measurement model was conceived in several steps. Firstly, the grounds of comparison had to be specified. Only the data manipulation part of the whole Big Data systems was selected as the their scope. The security aspect, quality assurance, and other non-functional requirements do not fall under the their scope. The same applies to the end-user decision making regarding the received final data and type specification of the data that enter the architecture. As to the actual model creation process, the initial requirement of all Big Data architectures was defined in the highest level as possible. Specifically, the general requirement for each architecture, which can be later extended or used only partially, is similar to the basic retrieve, transform, and consume process. The process starts with data retrieval from various sources, then the data are processed and transformed, and finally, the result of the manipulation is returned. This simple principle was then extended in every aspect within selected grounds for comparison, and a process of data modification and preservation was applied to each step. Afterward, this measurement model was validated against available architectures both from the scientific domain and architectures used in industrial applications. A few changes were applied to extend the model to cover more business and theoretical scenarios of Big Data capabilities. <Figure 1> is a result of the analysis process and should cover most of the scenarios required for big data architectures in the set their scope.

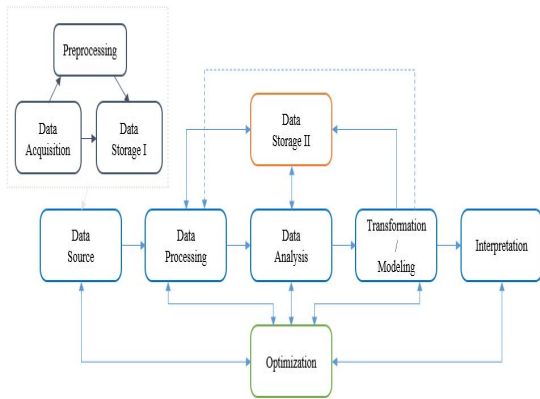


Figure 1. Measurement Model for BDP

5. Social Commerce for BDP

Social commerce architecture was built by analyzing the existing architectures from the industry domain, namely Naver, Instagram, Twitter, and so on, Network traffic measurement. All these existing applications were then mapped to the their architecture to prove the universality of the proposed architecture. The high-level design of the resulting architecture can be seen in <Figure 2> Data sources, Data extraction, Data loading and pre-processing, Data processing, Data analysis, Data loading and transformation, Interfacing and visualization, Data Storage, and Job and model specification. Data sources can be categorized into two dimensions: mobility and structure of data. Data sources refer only to the actual data and do not describe the data retrieval, that is the purpose of the Data extraction component. These two components of the architecture correspond with the Data acquisition component of the conceptual measurement model.

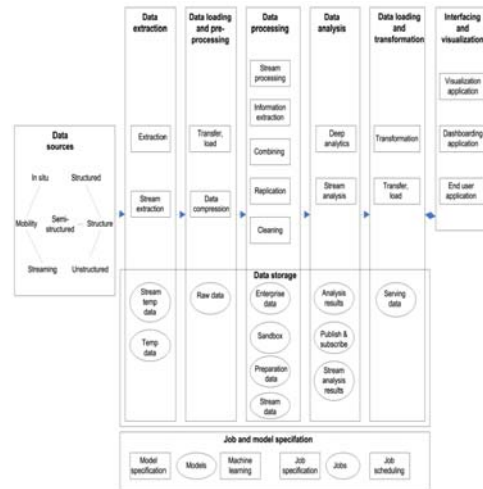


Figure 2. Social Commerce for BDP

Consecutively, the raw data can be transformed, loaded, or compressed in the Data-loading and preprocessing component. This raw data transformation matches the Pre-processing component of the measurement model. The next component of the architecture can also transform data in terms of combining, cleaning, or replication. Furthermore, the stream processing or information extraction can be defined in the scope of Data processing component, which has the same name in the measurement model and can be mapped directly. Similarly, the Data analysis maps to the Data analysis component, where deep analytics or stream analytics take place. The following step can be an additional transformation of the results of data analysis, which takes place in the Data loading and transformation component. The counterpart of this component in the measurement model is Transformation and modeling. It also includes the model specification of the architecture, which is defined separately in

the Job and model specification as a standalone process. Visualization of the analyzed and transformed data is part of the Interfacing and visualization component, which is a correlative component to the Interpretation part of the measurement model. Data storage component is spanned across all the components described above and resembles the previous architecture, where the data storage solution was very similar. The same way as in the previous example, this is represented in the conceptual measurement model in two parts as primary and secondary Data storage.

6. Conclusions

In this paper, we have proposed a description of the theoretical background on the Big Data. The main definition and characteristics of Big Data are described together with Hadoop and its specifications in terms of the main parts which are HDFS, MapReduce, and Common. contains the definition of the measurement model provided for the evaluation of Big Data architectures. The model is structured as multiple interacting components, where almost all the components can be treated as optional. Each component is described with expected functionality and the scope of responsibilities. The our scope for this conceptual measurement model was set with the emphasis on the data flow process and its orchestration, not on the data itself nor the attributes of the system. Then selected BDP architecture from both conceptual and industry origin were evaluated and mapped to the

measurement model for a unified point of comparison. Each evaluated architecture contains a description of the process used in the architecture. Our BDP architecture was randomly selected with the emphasis on the relevance to the Big Data architecture domain, and we shown <Figure 3>.

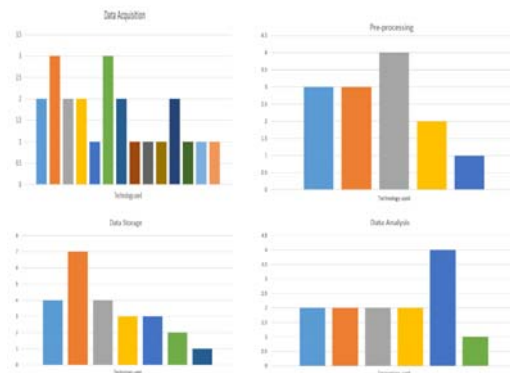


Figure 3. The Result of Social Commerce of Each Big Data Platform

It was possible to map all architectures to the conceptual measurement model allowing further evaluation in terms of technologies. This measurement model is not meant as the architecture proposal, but a unified structure on which was later possible to conduct group analysis for all different architectures with different approaches to the Big Data problematics. The architectures were similar in many ways, but some of them used various and unique approaches. Also, the purpose of each architecture differed based on the domain to which it was proposed.

References

- [1] P. Gaikwad, A. Mandal, P. Ruth, G. Juve, D. Król, and E. Deelman. *Anomaly detection for scientific workflow applications on networked clouds*. In: High Performance Computing & Simulation (HPCS), 2016 International Conference on. IEEE, pp. 645-652, 2016.
- [2] N. Pandeewari, and G. Kumar. *Anomaly Detection System in Cloud Environment Using Fuzzy Clustering Based ANN*. In: Mob. Netw. Appl. 21.3, pp. 494-505, 2016.
- [3] R. J. Hyndman, E. Wang, and N. Laptev. *Large-scale unusual time series detection*. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE, pp. 1616-1619, 2015.
- [4] B. Agrawal, A. Chakravorty, C. Rong, and T. W. Wlodarczyk. *R2Time: A Framework to Analyse Open TSDB Time-Series Data in HBase*. In: Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference, pp. 970-975, 2014.
- [5] M. Solaimani, M. Iftekhhar, L. Khan, B. Thuraisingham, and J. B. Ingram. *Spark-based anomaly detection over multi-source VMware performance data in real-time*. In: Computational Intelligence in Cyber Security (CICS), 2014 IEEE Symposium, pp. 1-8, 2014.
- [6] Y. Chen, J. Qian, and V. Saligrama. *A new one-class SVM for anomaly detection*. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference, pp. 3567-3571, 2013.
- [7] L. Yu and Z. Lan. *A scalable, non-parametric anomaly detection framework for hadoop*. In: Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference, pp. 22, 2013.
- [8] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica. *Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing*. In: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, p. 2, 2012.
- [9] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. *Rank-sparsity incoherence for matrix decomposition*. In: SIAM Journal on Optimization 21.2, pp. 572-596, 2011.
- [10] C. Wang, K. Viswanathan, L. Choudur, V. Talwar, W. Satterfield, and K. Schwan. *Statistical techniques for online anomaly detection in data centers*. In: Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on. IEEE, pp. 385-392, 2011.
- [11] J. S. Kim, J. H. Kim, *A Study on Adaptive Smart Platform for Intelligent Software in Big Data Environment*, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 15, No. 3, pp. 347-355, Jun. 2020.
- [12] C. Wang, V. Talwar, K. Schwan, and P. Ranganathan. *Online detection of utility cloud anomalies using metric distributions*. In: Network Operations and Management Symposium (NOMS), 2010 IEEE, pp. 96-103, 2010.
- [13] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*. In: Advances in neural information processing

systems, pp. 2080-2088, 2009.

[14] K. H. Ramah, H. Ayari, and F. Kamoun. *Traffic anomaly detection and characterization in the tunisian national university network*. In: NETWORKING 2006. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems. Springer, pp. 136-147, 2006

[15] M. Jahrer, A. Toscher, J. Y. Lee, J. Deng, H. Zhang, and J. Spoelstra. *Ensemble of collaborative filtering and feature engineered models for click through rate prediction*. In: KDDCup Workshop. 2012.

[16] D. E. Bloom, D. Canning, and A. Lubet. *Global population aging: Facts, challenges, solutions & perspectives*. In: Daedalus, 144.2, pp. 80-92, 2015.

소셜커머스 비교 분석을 이용한 빅데이터 플랫폼 아키텍처 기반 개념 측정 모델에 관한 연구

황성태

한성대학교 공과대학 조교수

요 약

빅데이터의 인기가 높아지면서 점점 더 많은 아키텍처가 제안되고 있지만 통합 솔루션은 없다. 본 논문의 주요 목적은 빅데이터 플랫폼 아키텍처를 매핑하고 비교할 개념적 측정 모델을 지정하고 준비하는 것이다. 개념 측정 모델은 책임 사양과 함께 여러 구성 요소로 분할되고 개념 측정 모델의 범위가 설정된다. 그런 다음 단일 구조로 매핑 된 모든 아키텍처를 평가하고 비교하고자 한다. 우리의 연구에는 빅데이터 개념에 대한 간략한 개요와 평가된 아키텍처에 사용되는 기술 요약이 포함되고 있다. 또한, 소셜커머스는 웹 3.0 기술과 소셜 데이터를 사용하여 교환 활동을

지원하는 상업 시장의 새로운 패러다임 인프라 모델이다. 상업적 시장에서의 하위 집합은 인기가 있었지만, 2012년 이래로 엄청나게 증가하였다. 특히, 일반적인 소셜커머스 관행을 넘어선 영역에서 프레임 워크와 애플리케이션의 효과에 대한 일반적인 연구가 진행 중이다. 본 연구는 4 가지 핵심 구성 요소가 있는 빅데이터 환경에서 포괄적인 소셜커머스 프레임워크를 개발하고자 한다. 그런 다음 빅데이터 플랫폼(BDP)을 적용하여 소셜커머스 성공 기업에 적용해 봄으로써 프레임 워크의 사용성과 적응성을 분석하고자 한다. 이에 따라 소셜커머스 기반 빅데이터 플랫폼 설계를 위한 일련의 지표를 제공하고자 한다. 제안된 프레임 워크는 소셜커머스 환경에서 설계 및 평가를 주도하는 데 사용된다. 본 연구는 새로운 디자인 프레임 워크를 제안하는데 있어 소셜커머스에 기여하고 있으며, 그들의 소셜 비즈니스에 대한 효과적인 접근을 위해 내부 제공에 대해서도 검토하고자 한다.

Acknowledgments

This research was financially supported by Hansung University.



Sung-Tae Hwang he is Assistant Professor of Department of IT Engineering at the Hansung University, Seoul, and Korea. He respectively received his Ph.D. degrees in Physics from Sogang University, Korea, in 2010. He is currently a senior researcher at Sogang University's Institute of Basic Science and is studying big data and quantum optics. His research interests include Quantum optics, Quantum information, Smart Platform, Machine Learning, Big Data, Intelligent Software, and Artificial Intelligent. He is a life member of the KKITS.

E-mail address: sthwang@hansung.ac.kr