International Journal of Intelligent Information Technologies

Volume 14 • Issue 4 • October-December 2018 • ISSN: 1548-3657 • eISSN: 1548-3665

An official publication of the Information Resources Management Association

IGIP
IGI PUBLISHING
AN IMPRINT OF IGI GLOBAL
WWW.IGI-GLOBAL.COM

International Journal of Intelligent Information Technologies
Volume 14 • Issue 4 • October-December 2018

Towards a Service-Oriented Architecture for Knowledge Management in Big Data Era

Thang Le Dinh, Université du Québec à Trois-Rivières, Trois-Rivières, Canada

Thuong-Cang Phan, Can Tho University, Can Tho, Viet Nam

Trung Bui, Adobe Research, San Jose, USA

Manh Chien Vu, Université du Québec à Trois-Rivières, Trois-Rivières, Canada

ABSTRACT

Nowadays, big data is a revolution that transforms conventional enterprises into data-driven organizations in which knowledge discovered from big data will be integrated into traditional knowledge to improve decision-making and to facilitate organizational learning. Consequently, a major concern is how to evolve current knowledge management systems, which are confronted with a various and unprecedented amount of data, resulting from different data sources. Therefore, a new generation of knowledge management systems is required for exploring and exploiting big data as well as for facilitating the knowledge co-creation between the society and its business environment to foster innovation. This article proposes a service-oriented architecture for elaborating a new generation of big data-driven knowledge management systems to help enterprises to promote knowledge co-creation and to obtain more business value from big data. The proposed architecture is presented based on the principles of design science research and its evaluation uses the analytical evaluation method.

KEYWORDS

Big Data Analytics, Big Data, Knowledge Management System, Service-Oriented

INTRODUCTION

The development and use of Knowledge Management Systems (KMSs) are currently having a direct and dramatic impact on business decisions and processes in modern and networked organizations that are required to be more competitive to grasp more business opportunities (Alavi & Leidner, 2001). However, these KMSs are currently confronted with a various and unprecedented amount of data, resulting from different business and IT-based services, called "big data" (Chen, Chiang, & Storey, 2012).

Big data provides high-volume, high-velocity and high-variety information assets that lead to a revolution of transforming traditional organizations into knowledge-intensive ones, called data-driven organizations (DDOs). Consequently, knowledge discovered in DDOs needs to be translated from big data into organizational knowledge to aid managers in making decision and in improving performance (Chen et al., 2012). Despite the fact that big data research has recently gained rapid growth, there is a

DOI: 10.4018/IJIIT.2018100102

lack of frameworks and architectures that enable DDOs to capture the value of big data in a systematic manner, especially for promoting organizational learning (Olivo, García Guzmán, Colomo-Palacios, & Stantchev, 2016; X. Wang, White, & Chen, 2015). Indeed, one of the most important challenges for KMSs today is to be able to deal with big datasets that are required to be updated frequently or continuously. Therefore, a new generation of KMSs to handle effectively big data sources becomes an essential tool for organizations, notably DDOs.

In fact, most of the recent studies, which are related to the integration of big data and KMSs, have separately concentrated on specific aspects of knowledge management such as business intelligence and business analytics (Chen et al., 2012), data mining and knowledge discovery (Wu, Zhu, Wu, & Ding, 2014) as well as the construction of the total process of collaborative knowledge management and set up the overall framework of a collaborative management system for IT start-up companies based on knowledge flows (Lu, 2016). These studies have strongly focused on knowledge exploration but have not been fully supported knowledge exploitation yet. There were little attempts to take into account the impact of big data on the whole process of organizational knowledge management and the trend of service orientation (Kakabadse, Kakabadse, & Kouzmin, 2003).

Besides, the service science perspective always considers customers as value co-creators. Co-creation is carried out by individuals and the organizations within a unified network that bring the benefits to all stakeholders (Vargo & Lusch, 2008).

As a result, this research aims at proposing a novel service-oriented architecture for big data-driven knowledge management systems, hereafter called BDD-KMS architecture. This paper discusses the architecture for a new generation of BDD-KMS to help enterprises, especially SMEs (small and medium-sized enterprises) to promote knowledge co-creation and obtain more business value from big data.

The rest of the paper is structured as follows. Section 2 provides a summary of theoretical background. Section 3 describes the essentials of the research design and positions our paper with respect to the related work. Section 4 introduces the proposed architecture according to the design science research, including a set of constructs, a model and a method. Section 5 demonstrates a use case of the architecture. Section 6 outline conclusions and some ideas for further research.

THEORETICAL BACKGROUND

In this research context, data consists of traditional data and big data. Knowledge is constituted from knowledge objects, which are classified on the basis of their level of development, that is, as data, information, knowledge or wisdom (Bierly, Kessler, & Christensen, 2000). Knowledge management (KM) is defined as organizational activities related to knowledge artefacts in which a learning process has occurred, and intellectual capital is accumulated and developed. Knowledge management systems (KMS) represent a specific type of information systems applied to handle organizational knowledge (Alavi & Leidner, 2001), that includes activities such as knowledge capture, knowledge organization, knowledge transfer and knowledge application (Le Dinh, Rickenberg, Fill, & Breitner, 2015). The backbone of KMSs is the knowledge architecture, which is the application of information architecture to knowledge management that supports and enhances the KM activities.

Big data brings a great opportunity but also a big challenge to implement KMSs in DDOs. The importance of big data does not revolve around how much data organizations have today, but what business value they can distill from this data in the best manner and within the shortest response time. In the global competitive environment, organizations, which are able to effectively leverage big data through KMSs, can differentiate themselves from their rivals. However, there is a great challenge for DDOs in handling the immense volumes of data, which are being continuously generated on an hourly basis. Hence, a huge computing power for analytics, which is required to process the unprecedented input, creates significant barriers for organizations, especially SMEs, to harness effectively the business value of big data.

Conventional KMSs were not prepared well for storing, processing and analyzing big data because they mainly used traditional databases. A survey of current research projects related to big data indicated that most of them have had only focused on specific aspects of KM such as such as business intelligence and business analytics (Chen et al., 2012), data mining and knowledge discovery (Wu et al., 2014) as well as knowledge flows (Lu, 2016). Knowledge flow is an interactive process of delivering the attained knowledge from the knowledge sender to the knowledge receiver through various tools. Between two layers of the knowledge flow, while the first layer is the knowledge flow within the enterprise, Kakabadse et al. (2003) argued that the second layer is the knowledge flow between the enterprise and the external, including enterprise customer base, competitors, other enterprises and scientific research institutes.

There are some other studies that focus on other technical aspects such as methods for building knowledge bases (Rezgui and Meziane, 2005; Suchanek & Weikum, 2014), knowledge fusion based on the machine learning (Dong et al., 2014), deep learning (Yu, 2013), random walk inference (Lao, Mitchell, & Cohen, 2011), and real-time stream data processing. However, these studies did not provide a comprehensive architecture for supporting the complete process of knowledge development as well as the collaboration of their systems to support organizational learning (Olivo et al., 2016; X. Wang et al., 2015). Recently, Cerchione & Esposito (2017) made a content analysis of the literature on KMSs in SMEs and found that these studies focused only on specific IT-based tools and information sharing practices, not on knowledge management practices.

On the other hand, the service-oriented approaches such as the Cloud computing and software as a service (SaaS) were regarded as an agile, flexible, and powerful approach for implementing KMSs (Šaša & Krisper, 2010). For this reason, we propose a general architecture for BDD-KMSs based on service-oriented principles that supports the whole knowledge management process in a DDO in order to provide a unified way of working, learning and innovating in the big data era.

RESEARCH DESIGN

This research proposes a novel service-oriented architecture, which can be considered as a design artifact. The design science research (DSR) has been chosen to carry out this research that aims at developing a technology-based solution as an artefact to overcome a business challenge (Hevner, March, Park, & Ram, 2004). Accordingly, we use the DSR methodology which includes six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

Activity 1: Problem Identification and Motivation

Looking at the problem relevance, the business challenge addressed by this research is how to implement effectively KMSs in the big data era, which facilitates the whole process of organizational knowledge management, including knowledge exploration and knowledge exploitation. This approach can help enterprises, especially SMEs, to implement their BDD-KMS with limited resources.

Activity 2: Define the Objectives for a Solution

Concerning the research rigour, the presented study has the theoretical foundations in knowledge management theory as well as in information system theory. In order to support the whole process of organizational knowledge development, this research addresses two main objectives: to capture and unify knowledge discovered from diverse big data and traditional data sources, and to support the process of organizational knowledge management.

Activity 3: Design and Development

Regarding the design artefact, the proposed architecture includes a set of constructs, a model and a method (Hevner et al., 2004) based on the concept of knowledge objects. We adopt a hybrid KM model that supports both the views of a knowledge object: epistemology-oriented and ontology-oriented KM models (Le Dinh et al., 2015). The first views a knowledge object as an entity that can be deconstructed into discrete and relevant attributes; meanwhile, the second defines knowledge solely through their relationships with a constructed universe of discourse. Furthermore, the proposed architecture is constructed based on diverse systems and services at different layers of knowledge objects such as data, information, knowledge and understanding layers.

Activity 4: Demonstration

We demonstrate the application of the architecture for a BDD-KMS for enterprises in general and a use case for a Quebec wine association and its members in particular. This use case is designed not only to scan the business environment information of the wine production and trade, it can be used also for the implementation of a customer KMS in each member of the association. Regarding to the limited number of employees and the business seasonality, the system allows to identify, store, organize and exploit the knowledge within a member and between the members of the association, especially knowledge about customers, products and environments. Subsequently, business owners can have a better understanding of the wine and derived market and all other challenges (customer's behavior, international competition, global challenges).

Activity 5: Evaluation

With respect to the search process, we focus on the exploration, implementation, experimentation and iteration. We study the related work and trends related to big data and KMSs, then implement or experiment (or both) the different aspects of the architecture to determine the suitability for real-world applications. The cycle, which generates design alternatives and test alternatives against business requirements, has been repeated several times during the period of 12 months by the research team, its students and business partners. With this trial version, we observed that the artifact supported well to the problem and a real-time way, mostly in the environment scanning system. We also proposed some management tools for individual business and the association. At the end of this activity, we can decide to improve the effectiveness of the artifact.

Activity 6. Communication

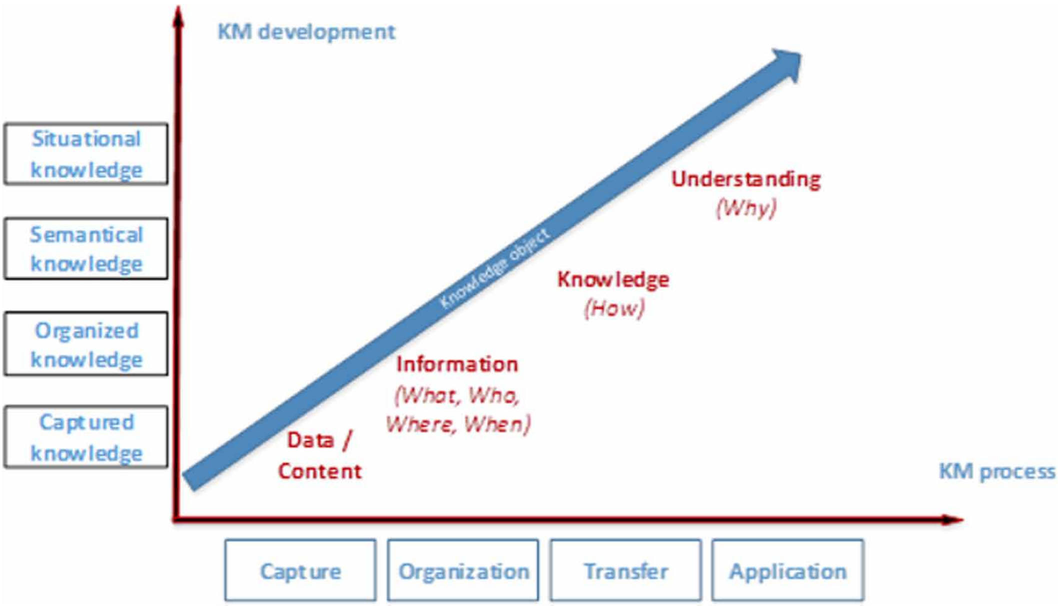
In respect of the design evaluation, the analytical evaluation method is used to validate the design artefact (Hevner et al., 2004). In particular, static analysis is used to examine the structure of the proposed architect, and the architecture analysis is used to study the fit of the design artefact into current information system architecture. The problem that we solve is not only for the wine sector, the artifact is effective at relevant SMEs. Our model is presented in some other public organizations who are in charges of local business development, and we have got the confirmation of their collaboration in our future experimentation of the proposed architecture.

SERVICE-ORIENTED ARCHITECTURE FOR BDD-KMSS

Constructs of the BDD-KMS Architecture

As mentioned above, constructs are different types of concepts related to knowledge objects and their knowledge components. As presented in Figure 1, "a knowledge object (KO) is a highly structured interrelated set of data or content, information, knowledge, and wisdom concerning some business situation", which provides a viable approach for dealing with the situation. In this study, the term "understanding" is used instead of the term "wisdom" since the research just focuses on the first

Figure 1. A knowledge object



level of understanding in which a DDO understands how to increase business values by using their knowledge and knowing. In other words, the proposed architecture is still at a level of knowledge-based decision support instead of the best decision-making support. According to the structure view, a knowledge management process covers both knowledge exploration and knowledge exploitation. The process of knowledge exploration concerns the capture and organization of different knowledge (tacit and explicit) in the organizational memory, whereas the process of knowledge exploitation concerns the transfer and application of classified and organized explicit knowledge in the knowledge repository (Le Dinh et al., 2015). A knowledge object is considered as a collection of knowing, called knowledge components (Le Dinh et al., 2015). Knowledge components can be stratified in different levels ranging from lowest to highest: cognitive (know-what), conditional (know-when), situational (know-where), applied (know-how) and rule-of-thumb (know-why). A knowledge component can be explicit or tacit knowledge as well as individual or collective knowledge. Knowledge components of knowledge objects can be used and shared by using different knowledge conversion processes such as socialization, externalization, combination and internalization.

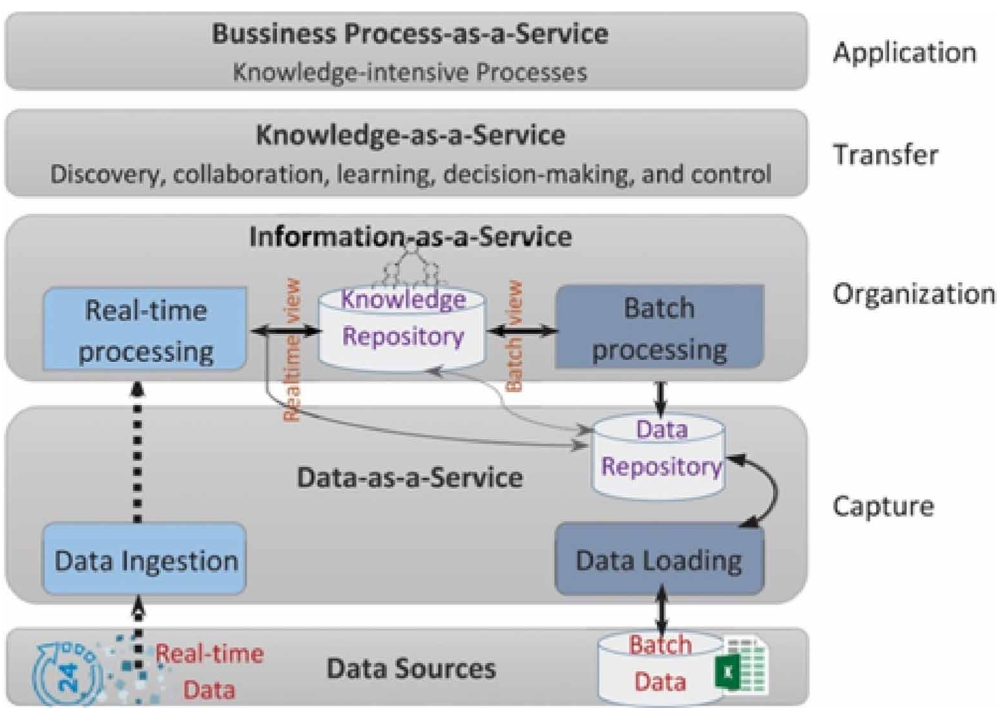
According to the behavior view, a life cycle of a knowledge object includes the following dynamic states: Captured knowledge, Organized knowledge, Semantic knowledge and Situational knowledge. Firstly, data is raw and simply exists in any form (Bellinger, Castro, & Mills, 2004). A KO is at the Captured knowledge state if its data is captured and stored in the knowledge repository. Secondly, information is data that has been given meaning by way of relational connection (Bellinger et al., 2004). A KO is at the Organized knowledge state if its data is organized according to its knowledge components corresponding to the structure of knowledge to become useful information. A KO at this level can answer the simple questions such as “What”, “Who”, “Where”, “When”. Thirdly, knowledge is the appropriate application of information to facilitate organizational activities (Bellinger et al., 2004). A KO is at the Semantic knowledge state if its knowledge components corresponding to the structure of knowledge are linked to the knowledge components corresponding to the transition of knowledge (Le Dinh et al., 2014). A KO at this level is able to provide an answer or guidelines for a “How” question. Fourthly, understanding is the process by which a person can take knowledge and synthesize new knowledge from the previous held knowledge to make business decisions (Bellinger et al., 2004). A KO is at the Situational knowledge state if its knowledge components corresponding to the structure and transition of knowledge are linked to the knowledge components corresponding to the coherence of knowledge (Le Dinh et al., 2014) so that it can provide guidelines for a person to take necessary actions faced a specific business situation.

Model of the BDD-KMS Architecture

A model is defined as the statements expressing the relationships among knowledge components of knowledge objects. Knowledge components can be classified based on the key aspects of knowledge that forms the knowledge structure of a DDO: structure, transition, and coherence of knowledge (Le Dinh et al., 2014). The structure of knowledge, which is represented by the "know-what" knowledge component and its associated knowledge-components (know-who, know-when, know-where), describes knowledge that relates to a phenomenon of interest (Garud, 1997). Know-what in a DDO can relate to products, services, or markets. The transition of knowledge, which is represented by the "know-how" knowledge component, describes the understanding of the generative processes that constitute phenomena (Garud, 1997). Know-how in a DDO concerns the organizational processes, which are a feedback of the organization to the occurrence of an event or a situation. The coherence of knowledge, which is represented by the "know-why" knowledge component, which describes the understanding of the principles underlying phenomena. Know-why in a DDO concerns the business rules and guidelines, which help to make business decisions face certain situations.

As presented in Figure 2, the service-oriented architecture consists of four layers: Data-as-a-Service, Information-as-a-Service (corresponding to the structure of knowledge), Knowledge-as-a-Service (corresponding to the transition of knowledge), and Business Process-as-a-Service layers (corresponding to the coherence of knowledge).

Figure 2. A Service-Oriented Architecture for Big Data-Driven KMSs



Data-as-a-Service layer (DaaS) includes data loading and ingestion components, a data repository, and data services. DaaS is a data collection, provision and distribution model in which big data is made available to higher-level layers and users over the network.

Information-as-a-Service layer (IaaS) contains batches and real-time processing components, a knowledge repository, and information services. IaaS structures and analyzes the disorder, disparate and obscure data from the batch and real-time data into linked, integrated, and semantic information that can be delivered on demand as a service. This information can be organized by tagging based on the corresponding knowledge components of knowledge objects. In addition, data files can be categorized and associated automatically or semi-automatically to a set of related concepts of the knowledge structure.

Knowledge-as-a-Service layer (KaaS) typically provides knowledge services supporting organizational activities such as knowledge discovery, collaboration, learning, decision-making, and control services (Alavi & Leidner, 2001). KaaS transfers the information from the IaaS layer into knowledge based on their semantic contexts and organizes this knowledge corresponding to the "know-how" knowledge component. Besides, it also provides the ability to analyze, learn, infer, and make the knowledge and experience available to the higher business layer and users.

Business Process-as-a-Service layer (BPaaS) includes process representation, which provides building blocks for aggregating the loosely-coupled services of the lower layers as a collaborative process aligned with business goals (Doloreux & Shearmur, 2012). The governance of the business processes has been supported by the "know-why" knowledge component that provides situational knowledge and guidelines for making decisions faced business situations (Le Dinh, Rinfret, Raymond, & Dong Thi, 2013).

Method of the BDD-KMS Architecture

A method is defined as the activities that support the process of organizational knowledge development. A BDD-KMS operates based on four main activities of the knowledge management process (Le Dinh et al., 2015): knowledge capture, organization, transfers and application. They are respectively realized by the DaaS, IaaS, KaaS and BPaaS layers as the following:

- The DaaS layer captures batches and real-time data through the data loading and ingestion components. The batch data is loaded, processed and saved to the data repository, whereas the real-time data is ingested and sent to the real-time processing components. The data services enable maximal mashup, reuse, and sharing of the available data in the data repository;
- Based on the knowledge structure, the batch and real-time data processing components of the IaaS layer filter, extract, analyze, integrate and migrate the batch and real-time data into batches and real-time information views, respectively. The information views are unified and saved to the knowledge repository that may store its data in the data repository. Besides, the real-time component also writes the processed real-time data to the data repository. The information services make the up-to-date information available to users;
- The KaaS layer is able to deduce knowledge from the information of IaaS through the knowledge services. The collaboration service allows the creation, sharing and application of the knowledge and users with support of tools. The discovery service provides functions such as search, retrieval, mining, mapping, navigation, and presentation of the knowledge. The learning service creates more knowledge for decision-making by enabling machine learning algorithms, data scientists, predictive modellers, and other analytic professionals. The decision-making service devises an increased understanding about a business task or agreed criteria, and a guideline to make the most effective and strategic business decision. The control services ensure the high-quality performance for business processes and their corresponding services. All the knowledge generated is then saved to the knowledge repository;

- The BPaaS layer, on top of the other service layers, delivers services to applications through combining them with knowledge-intensive business processes. The main purpose of this layer is to deal with the practices and applications to facilitate the development of intellectual capital by applying the knowledge repository to a special use or purpose. There are three primary mechanisms for knowledge applications: directives, organizational routines, and self-contained task teams (Le Dinh et al., 2015).

It is obvious that the proposed architecture satisfies the two objectives of the research question. Namely, the DaaS and IaaS layers carry out the capturing and unifying knowledge discovered from big data (based on the architecture's constructs and model). Moreover, all the layers are designed according to SOA principles to support the activities of the knowledge management process (based on the architecture's method). As a result, the architecture can take the benefits from SOA and enhance the added value from big data in order to promote the organizational knowledge management process.

Design Evaluation

The design evaluation is based on the analytical evaluation method (Hevner et al., 2004) to evaluate the utility, quality and efficacy of the proposed architecture. For the time being, this research focuses on the utility and quality of the architecture. Therefore, this evaluation has been carried out the following steps: exploration, implementation, experimentation and iteration.

The exploration step concerns the static analysis, which examines the artefacts for static qualities. This step has been performed by two researchers, one in information systems and one in knowledge management, who play the role of the users of the design. These users evaluate the design and then the artefacts of the design. The results of the first step have been used to design the architecture as mentioned in Figure 2 that supports all the knowledge development levels.

The implementation step concerns the architecture analysis, which studies the suitability of the design artefacts into current technologies and tools related to big data and information system architecture. The results of the second step helped us to fulfil the architecture, which is composed of important features and viable combination of software tools to meet the requirements of the BDD-KMS architecture. The result of this step is a prototype built by the research team, including the researchers, PhD and master's students.

The experimentation step concerns the experimentation with assuming users and real users. As mentioned above, the evaluation cycle is 12 months so that the team members can evaluate the prototype (design artefact) with their students during an academic year. In general, the two groups of undergraduate students (about 40 students per group) evaluate the information-as-a-service provided by the prototype and a group of MBA students (about 30 students per group) evaluate the knowledge-as-a-service. Firstly, students can be taught how to use the services at the lower level (data-as-a-service for undergraduate students and information-as-a-service for MBA students) in order to carry out their term projects, which design or elaborate a knowledge infrastructure for a SME. Secondly, students use those services to build, design, or suggest new services at the higher level (visualization, reporting and data analytics for MBA students, data processing and integration for undergraduate students). Thus, the research team evaluates the results of their work and reports as well as improves the design artefact based on the feedback of students. Lastly, the business partners such as some selected SMEs evaluate the design artefact during the summer time. In some cases (and also preferred cases), the students who already evaluated the artefact during the academic year, can evaluate it again during their internship at the business partner.

The iteration step concerns the validation of the design artefact to see whether the artefact meets the business requirements or not. In order to determine the best-fit architecture for BDD-KMS for SMEs, the research team has performed the three cycles: the first cycle focusing on NoSQL databases, the second cycle focusing on Spark, and the third cycle focusing on Cloud-native Spark.

The first cycle was the validation of a prototype based on the ecosystem of Neo4J (neo4j.com), a NoSQL database management system, and Python. This ecosystem has limited functionalities and therefore, it was not easy for students to experiment and create new services.

The second cycle concerned the validation of a prototype-based Hadoop & Spark framework. This prototype has a really good ecosystem and rich functionalities. Students can use different tools and systems to work with their services. However, such complex infrastructure is still a hard challenge for an enterprise, especially SMEs, since it required a good knowledge about the IT infrastructure.

The third cycle used the Cloud computing services for implementing the Spark ecosystem. Indeed, new services provided by Amazon, Google, and Microsoft help enterprises to set up and manage a robust and complex infrastructure such as Hadoop and Spark. This cloud-native ecosystem can be integrated easily with end-user tools so that users can create new services and co-create more value with the BDD-KMS system.

USE CASE

This use case is extracted from a BDD-KMS architecture for a Quebec wine association that aims at performing business environment scanning for the association and at improving the customer service for its members based on the knowledge about business environment and customers.

Environment scanning is defined as the process of monitoring and acquiring information about events and their relationships in a company's internal and external environments. The impressive expansion of the Internet, social networks, mobile and media technologies have generated a large number of digital data available to firms. Meanwhile, environmental scanning practices are generally a real challenge for SMEs since they lack sufficient resources and capacities to set up a formal system to conduct environmental scanning or typically lack the necessary infrastructure to search for and gather information in a suitable manner (Yoo & Sawyerr, 2014).

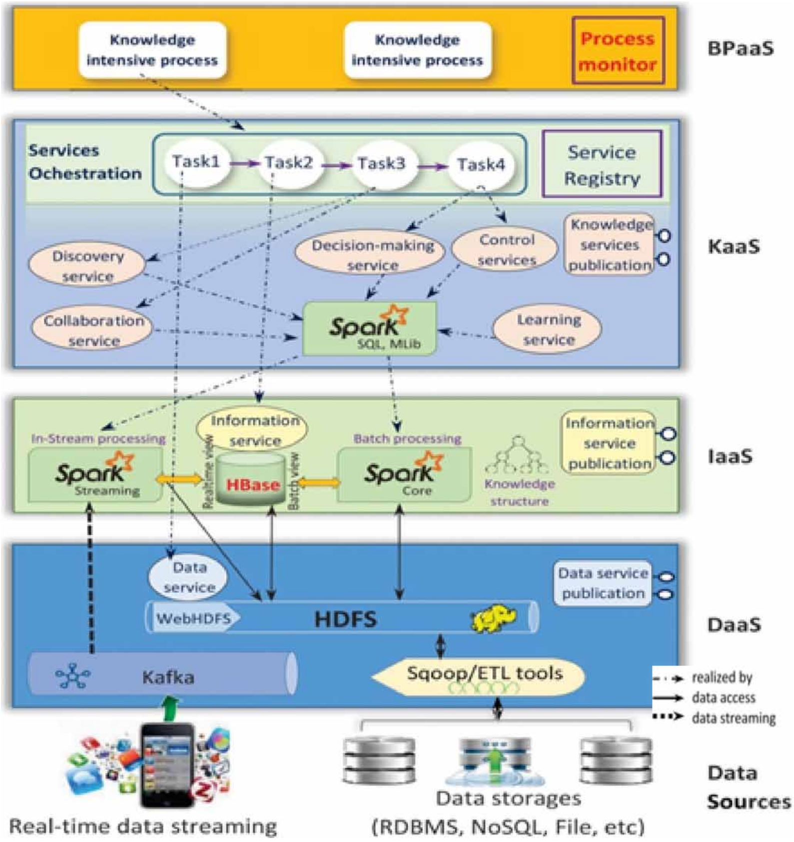
A customer KMS helps SMEs to renovate, conceive ideas, exploit the thinking potentials of customers, and fix their weaknesses due to lack of experiences in management and production (Rababah, Mohd, & Ibrahim, 2011). However, SMEs has still not really attached special importance to and invested in customer KMSs nor taken to follow advantage of the information from big data and social network to manage and expand sources of customers. It is this that has resulted in a lot of restrictions in developing the full potential of SMEs. In practice, there is a large number of SMEs still managing information according to the mechanism for collecting, analyzing information, particularly customer information in a classical way which takes managers much time when making decisions, thus increasing risk and reducing the possibility to optimize product and service quality (Oparaocha, 2015).

Figure 3 presents the BDD-KMS architecture for a Quebec wine association. Data from different sources are extracted, organized and transformed into information and knowledge to develop environment and customer insight. Several big data technologies have been launched and could replace highly customized, expensive legacy systems with a standard solution that is able to run on commodity hardware or the cloud environment. Most often, the open-source technologies are prominent candidates to be used in big data projects because they can be implemented far less expensive than proprietary technologies. Besides, these technologies have a great support community as well.

Technical specifications for implementing the service layers of the BDD-KMSs are described as follows:

- **The Data Sources Layer:** The batch data sources come from files, databases, or information systems. In case of this demontrstaion, it is recommended that the members of associations implement and use the open-source systems such as Wordpress (wordpress.org) for E-commerce and Enterprise Content Management, SugarCRM (sugarcrm.com) for Customer Relationship Management and Odoo (odoo.com) for Enterprise Resource Planning systems. Besides, relational DBMS or NoSQL-based systems can be used for storage and retrieval of data sources such as

Figure 3. A Use Case of a Service-Oriented BDD-KMS



log files and click-through of web applications and feeds, events, comments, messages, and images of social networks;

- **The DaaS Layer:** Hadoop Distributed File System (HDFS) is selected as the data repository of the system. It is a well-known fault-tolerant distributed file system of Apache Hadoop (hadoop.apache.org), a dominant framework for big data processing with the large infrastructure being deployed and used in manifold application fields. Concerning data capture tools, Apache Sqoop (sqoop.apache.org) is widely used for efficiently exporting or importing bulk data between HDFS and structured data storage such as relational databases. ETL (Extract, Transform, and Load) tools can be used to bulk load data from files or NoSQL to HDFS. On the other hand, Apache Kafka (kafka.apache.org), which is a distributed, reliable, high-throughput and low latency publish-subscribe messaging system, is used for stream processing. The data service is executed by WebHDFS for Hadoop that enables users and systems to access data in HDFS using an industry-standard mechanism;
- **The IaaS Layer:** The IaaS is implemented based on Cloud native Apache Hbase and Spark. Apache HBase is a distributed column-oriented NoSQL database built on top of HDFS that supports random, real-time read/write access to HDFS, and the bulk loading feature. Since HBase can provide web service gateways, this technology has been used to implement the information services related to the knowledge repository. Unlike a conventional KMS architecture designed for structured and internal data, the big data-driven KMS architecture works with raw unstructured and structured data as well as internal and external data sources. Since the ability

to handle batch and (near) real-time data is required, Apache Spark is selected for both the batch and real-time data processing. Spark has emerged as the next-generation big data processing engine because it works with data in memory that is faster and better to support a variety of compute-intensive tasks (Shanahan & Dai, 2015). Notably, Spark's components, including stream processing (Streaming), machine learning library (MLlib), structured data querying (SQL), and GraphX graph processing, are built on the same core architecture (Core/Engine) for distributed analytics. Spark Core works with the batch data from the data repository (HDFS) to organize content according to their semantics and to create and maintain the knowledge base. Indeed, according to service science perspective, customers are considered as co-producers or value co-creators. Therefore, the knowledge base contains the knowledge for customers (focusing on knowledge about products and services), knowledge from customers (focusing on the interaction between customers, the organization and social networks), and knowledge about the customer (focusing on customer profiles, segmentation, and activity history). Concerning the business environment, the knowledge base also concerns the knowledge about technologies, partners, suppliers, competitors, policies and laws;

- **The KaaS Layer:** From a technical perspective, the system built upon SOA (Service-oriented architecture) principles is constituted from the services as mentioned earlier. They are defined by a description language with callable interfaces to support business processes and are implemented in different programming languages. As a result, we found an appropriate technology, Web services (WS), including RESTful Web services, suiting to the implementation of the knowledge services. Web services are the most popular and well-known technology to implement SOA, both in the industry and the academia (Chollet & Lalanda, 2011). Knowledge objects are built based on customers, products, services, technologies, partners, suppliers, competitors, policies and laws (know-what), activities (know-how) and understanding of the situation related to customers, suppliers, products, services and their activities (know-why). Services at this layer aim at creating a new knowledge object, modifying properties and relationships of a knowledge object, or consulting content resources of knowledge objects;
- **The BPaaS Layer:** The primary mechanisms for knowledge applications can be directives and organizational routines. Directives refer to the specific set of rules developed through the conversion of tacit knowledge to explicit knowledge for efficient; meanwhile, organizational routines refer to the development of process specifications that allows individuals to apply and integrate their specialized knowledge without the need to communicate what they know to others (Le Dinh et al., 2013). Thanks to the experimentation with MBA students, a set of scorecards and KPIs (Key Performance Index) have been developed to help enterprises to build their directives. For the time being, the use case focuses on the business processes such as business development, and sales and marketing process and then to use the workflow application to support the automatization of organizational routines (Minor, Tartakovski, & Schmalen, 2008), which is based on the services at the BPaaS layer such as decision-making and control services. Besides, the rationalization of organizational routines as well as the integration of directives into processes to facilitate organizational learning and intelligent capital development can be carried out thanks to the knowledge discovery, collaboration, and learning services. For instance, the knowledge discovery service allows to ask a what-question to find information related to a knowledge object, to ask a how-question to know how to use a function or to carry out an activity, or to ask a why-question to understand a particular situation.

CONCLUSION

In the era of big data, one of the most important characteristics of knowledge management systems (KMSs) should be big data-driven to leverage all available data as a competitive advantage. We presented a service-oriented architecture for implementing big-data driven KMSs (BDD-KMS).

According to our knowledge, it is one of the first service-oriented architectures that concentrates on reconciliation of big data and KMSs to facilitate organizational learning and co-creation process based on the concept of knowledge objects. Compared with the related work such as Cerchione & Esposito (2017), X. Wang et al. (2015), Olivo et al. (2016), our approach not only concentrates on the IT-based level but also on the whole knowledge management process.

Our approach aims at adding more business value from big data and at facilitating knowledge development, co-creation, and organizational learning. Big data has transformed enterprises into data-driven organizations, which require the foundation to transform data into knowledge, optimize decisions, and maximize profits. The approach helps data-driven organizations to build a new-generation of KMSs based on the service-oriented perspective that supports the organizational knowledge development process, and the unification of knowledge derived from diverse (big) data sources. By applying service-oriented principles, an organization can manage and govern business and digital transformation, setting them apart from their competitors. The benefits include seamless integration, cloud enabled solutions, holistic business insight and organizational agility. In particular, the approach can help SMEs, which have limited resources, to develop their own BDD-KMS architecture based on open-source systems and cloud-based services to obtain more business value from big data.

The evaluation of the proposed architecture is still limited in the academic environment, we intend to perform other analytical design evaluations such as optimizational and dynamic analysis to evaluate the efficacy of the architecture in the future. Currently, an ongoing project is being carried out to deploy an architecture for the Quebec wine associations and some selected members for business environment scanning and customer knowledge management. This project uses the BDD-KMS approach to build a BDD-KMS based on Cloud-native Spark and then integrates this system with other information systems and services to capture and organize environment and customer knowledge in a coherent way. After this evaluation, we believe that the proposed architecture can be used widely to develop and deploy the BDD-KMSs in practice.

Finally, there are some critical research directions to increase the business value of our BDD-KMS approach: A workflow engine for autonomous and run-time execution of business processes in the BPaaS layer and user-centric context aware services co-operating with this workflow engine.

REFERENCES

Alavi, M., & Leidner, D. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *Management Information Systems Quarterly*, *25*(1), 107–136. doi:10.2307/3250961

Bellinger, G., Castro, D., & Mills, A. (2004). Data, Information, Knowledge, and Wisdom. Retrieved from http://www.systems-thinking.org/dikw/dikw.htm

Bierly, P. E. III, Kessler, E. H., & Christensen, E. W. (2000). Organizational learning, knowledge and wisdom. *Journal of Organizational Change Management*, *13*(6), 595–618. doi:10.1108/09534810010378605

Cerchione, R., & Esposito, E. (2017). Using knowledge management systems: A taxonomy of SME strategies. *International Journal of Information Management*, *37*(1, Part B), 1551–1562. doi:10.1016/j.ijinfomgt.2016.10.007

Chen, H. C., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: from big data to big impact. *Management Information Systems Quarterly*, *36*(4), 1165–1188.

Chollet, S., & Lalanda, P. (2011). A Model-Driven Approach to Service Composition with Security Properties. In J. Lee, S.-P. Ma, & A. Liu (Eds.), *Service Life Cycle Tools and Technologies: Methods, Trends and Advances* (pp. 154–174). IGI Global.

Doloreux, D., & Shearmur, R. (2012). Collaboration, information and the geography of innovation in knowledge intensive business services. *Journal of Economic Geography*, *12*(1), 79–105. doi:10.1093/jeg/lbr003

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., . . . Zhang, W. (2014). Knowledge vault: a web-scale approach to probabilistic knowledge fusion. *Paper presented at the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/2623330.2623623

Garud, R. (1997). On the distinction between know-how, know-why and know-what. *Advances in Strategic Management*, *14*, 81–101.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in Information Systems research. *Management Information Systems Quarterly*, *28*(1), 75–105. doi:10.2307/25148625

Kakabadse, N. K., Kakabadse, A., & Kouzmin, A. (2003). Reviewing the knowledge management literature: Towards a taxonomy. *Journal of Knowledge Management*, *7*(4), 75–91. doi:10.1108/13673270310492967

Lao, N., Mitchell, T., & Cohen, W. W. (2011). Random walk inference and learning in a large scale knowledge base. *Paper presented at the Conference on Empirical Methods in Natural Language Processing*.

Le Dinh, T., Ho Van, T., & Moreau, E. 2014. A Knowledge Management framework for Knowledge-Intensive SMEs. In *Proceeding of 16th International Conference on Enterprise Information Systems* (pp. 435-440).

Le Dinh, T., Rickenberg, T., Fill, H. G., & Breitner, M. (2015). Enterprise Content Management Systems as a Knowledge Infrastructure: The Knowledge-Based Content Management Framework. *International Journal of e-Collaboration*, *11*(3), 49–70. doi:10.4018/ijec.2015070104

Le Dinh, T., Rinfret, L., Raymond, L., & Dong Thi, B.-T. (2013). Towards the Reconciliation of Knowledge Management and e-Collaboration Systems. *Interactive Technology and Smart Education*, *10*(2), 95–115. doi:10.1108/ITSE-09-2012-0022

Lu, J. (2016). Research on Enterprise Knowledge Management System based on Computer Big Data Processing. *Revista Ibérica de Sistemas e Tecnologias de Informação,* (E7), 281-293.

Minor, M., Tartakovski, A., & Schmalen, D. (2008). Agile Workflow Technology and Case-Based Change Reuse for Long-Term Processes. *International Journal of Intelligent Information Technologies*, *4*(1), 80–98. doi:10.4018/jiit.2008010104

Olivo, L. J. F., García Guzmán, J., Colomo-Palacios, R., & Stantchev, V. (2016). IT innovation strategy: Managing the implementation communication and its generated knowledge through the use of an ICT tool. *Journal of Knowledge Management*, *20*(3), 512–533. doi:10.1108/JKM-06-2015-0217

Oparaocha, G. O. (2015). SMEs and international entrepreneurship: An institutional network perspective. *International Business Review*, *24*(5), 861–873. doi:10.1016/j.ibusrev.2015.03.007

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. doi:10.2753/MIS0742-1222240302

Rababah, K., Mohd, H., & Ibrahim, H. (2011). Customer Relationship Management (CRM) Process from Theory to Practice: The Preimplementation Plan of CRM System. *International Journal of e-Education, e-Business, e-management and e-learning, 1*(1).

Rezgui, Y., & Meziane, F. (2005). A Web Services Implementation of a User-Centered Knowledge Management Platform for the Construction Industry. *International Journal of Intelligent Information Technologies*, *1*(4), 1–19. doi:10.4018/jiit.2005100101

Šaša, A., & Krisper, M. (2010). Knowledge management in service-oriented systems. In *Proceedings of the 2010 conference on Information Modelling and Knowledge Bases XXI* (Vol. 206, pp. 89-104).

Shanahan, J. G., & Dai, L. (2015). Large Scale Distributed Data Science Using Apache Spark. *Paper presented at the 21th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia. doi:10.1145/2783258.2789993

Suchanek, F. M., & Weikum, G. (2014). Knowledge bases in the age of big data analytics. *Proc. VLDB Endow., 7*(13), 1713-1714. doi:10.14778/2733004.2733069

Vargo, S., & Lusch, R. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of Marketing Science*, *36*(1). doi:10.1007/s11747-007-0069-6

Wang, X., White, L., & Chen, X. (2015). Big data research for the knowledge economy: Past, present, and future. *Industrial Management & Data Systems*, *115*(9), IMDS-09-2015-0388. doi:10.1108/IMDS-09-2015-0388

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *Knowledge and Data Engineering*. *IEEE Transactions on*, *26*(1), 97–107. doi:10.1109/TKDE.2013.109

Yoo, S.-J., & Sawyerr, O. (2014). The Search for Broad Environmental Information and New Product Development Success in Technology-Based SMEs. *Journal of Enterprising Culture*, *22*(01), 27–55. doi:10.1142/S0218495814500022

Yu, K. (2013). Large-scale deep learning at Baidu. *Paper presented at the 22nd ACM International Conference on Conference on Information and Knowledge Management*.

Thang Le Dinh is associate professor in Information Systems at the Université du Québec à Trois-Rivières, Canada. He is also director of the DIMALAB Living Lab for promoting the digital transformation of marketing (dimalab.ca) and co-director of the LARIDEPED laboratory (larideped.org), whose purpose is to promote business development in the digital age. His primary research interest focuses primarily on Service modeling, Knowledge management, Enterprise systems, E-collaboration, Project management and smart services. His recent research focuses on the management, the science and the engineering of smart services, in particular, an integrated framework that helps knowledge-intensive enterprises in building a knowledge infrastructure efficiently to support smart services.

Thuong-Cang Phan has been a senior lecturer of Computer Networks Department, College of Information Technology, Can Tho University, Vietnam since 1998. He received a master's degree in computer science from Asian Institute of Technology, Thailand, in 2006. After 2.8 years he finally got his PhD degree in computer science from Blaise Pascal University (Université Clermont Auvergne), France in 2014. I worked as a post-doc researcher in Université du Québec à Trois-Rivières, Canada in 2015. His research interests include (big) data algorithms and architectures, service-oriented information systems, big data-driven knowledge management systems. He has published papers in TLDKS (Trans. Large-Scale Data- and Knowledge-Centered Systems), AMCIS (Americas Conference on Information Systems), IESS (International Conference on Exploring Service Science), ACIIDS (Asian Conference on Intelligent Information and Database Systems), IUKM (Integrated Uncertainty in Knowledge Modelling and Decision Making), etc. He served as an associate editor in ECIS 2018 (European Conference on Information Systems) – Track: Knowledge-intensive smart services and their applications. He has been involved in research projects such as the STIC ASIA GOD project (30 000 euros funding, 2013-2015), Vietnam's National project.

Trung Bui is a researcher at Creative Intelligence Lab, Adobe Research. He develops machine learning algorithms for Adobe digital media and digital marketing products. Broadly, he is interested in applying AI techniques to intelligent applications such as personal assistant agents, intelligent tutoring, spoken dialogue systems, multimodal interfaces, conversational video games. He joined Adobe Research in 2012. Before that, he was at North Side Inc. in Montreal where he led the language team to develop dialog management technologies for the Bot Colony game. He conducted his PhD research at Human Media Interaction Group, University of Twente, the Netherlands and worked as an engineering research associate at Stanford University in the Center for the Study of Language and Information. He also worked at the Artificial Intelligence Laboratory, Swiss Federal Institute of Technology in Lausanne, Switzerland.

Manh Chien Vu is currently postdoctoral researcher at the Université du Québec à Trois-Rivières, Canada. He was professor in finance at the Thuong Mai university, Vietnam. His research interests include knowledge management, business environment scanning, service science, SMEs, and performance management systems.