Industrial Marketing Management xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Industrial Marketing Management

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

Real-time big data processing for instantaneous marketing decisions: A problematization approach

Abdul Jabbar[a,*], Pervaiz Akhtar[b], Samir Dani[a]

[a] Huddersfield Business School, University of Huddersfield, UK
[b] Kent Business School, University of Kent, Canterbury, UK

ARTICLE INFO

Keywords:
Real-time processing
Batch processing
Internet of things
Social media
Programmatic marketing
Decision making
Big data
Problematization

ABSTRACT

The collection of big data from different sources such as the internet of things, social media and search engines has created significant opportunities for business-to-business (B2B) industrial marketing organizations to take an analytical view in developing programmatic marketing approaches for online display advertising. Cleansing, processing and analyzing of such large datasets create challenges for marketing organizations — particularly for real-time decision making and comparative implications. Importantly, there is limited research for such interplays. By utilizing a problematization approach, this paper contributes through the exploration of links between big data, programmatic marketing and real-time processing and relevant decision making for B2B industrial marketing organizations that depend on big data-driven marketing or big data-savvy managers. This exploration subsequently encompasses appropriate big data sources and effective batch and real-time processing linked with structured and unstructured datasets that influence relative processing techniques. Consequently, along with directions for future research, the paper develops interdisciplinary dialogues that overlay computer-engineering frameworks such as Apache Storm and Hadoop within B2B marketing viewpoints and their implications for contemporary marketing practices.

1. Introduction

We embed our contributions in big data processing in the context of business-to-business (B2B) industrial marketing, specifically the targeting of real-time online display advertising linked with programmatic marketing. To do that, we adopt a theory blending approach (Oswick, Fleming, & Hanlon, 2011), drawing on concepts from both B2B programmatic marketing (Li, Ni, Yuan, & Wang, 2018; Mouzas & Araujo, 2000) and big data processing literature (Ahmed et al., 2017; Cao, Duan, & El Banna, 2018; van Dieijen, Borah, Tellis, & Franses, 2019). Big data is fundamentally, *extremely large datasets, made up of structured and unstructured data that can be processed and analyzed to reveal patterns and trends* (Hazen, Boone, Ezell, & Jones-Farmer, 2014) and programmatic marketing *as the automated purchase of marketing material, content or advertising space based on current user sessions and experiences* (McGuigan, 2019; Stevenson, 2015). Extant literature in the field of marketing suggests that programmatic marketing is a relatively new concept in technological development terms. However, it has emerged as a stimulus for organizational marketing decision making — contributing to business value creation by improving operational, social, environmental and financial performance (Akhtar, Frynas, Mellahi, & Ullah, 2019) and has generated much interest in both academic and business communities (H. Chen, Storey, Chiang, & Storey, 2012). In a similar vein there is an evolution taking place in marketing, moving away from traditional approaches, towards online dynamic, targeted and analytical marketing decision making (Li et al., 2018). However, by looking at the nexus of big data analytics and programmatic marketing we identify further opportunities for strengthening the discussion in B2B marketing (see Table 1). Therefore, this study seeks to extend current literature in big data analytics and programmatic marketing in order to develop a deeper understanding of the role of data in real-time online display advertising for B2B markets.

As industrial marketing organizations continue to migrate online for their marketing requirements, the need for big data analytics in terms of speed, efficiency and accuracy is increasing (Cao et al., 2018; Xu, Frankwick, & Ramirez, 2016). In addition, the changing nature of data, the amalgamation of structured and unstructured datasets will create increasing opportunities in a variety of fields and contexts (Akhtar, Tse, Khan, & Rao-Nicholson, 2016; Caputo, Marzi, & Pellegrini, 2016). There are however several caveats, obstacles and challenges, specifically with relation to issues around data storage, software development and infrastructure (Hashem et al., 2015) which remain unaddressed,

* Corresponding author.
E-mail addresses: a.jabbar@hud.ac.uk (A. Jabbar), S.S.Dani@hud.ac.uk (S. Dani).

https://doi.org/10.1016/j.indmarman.2019.09.001
Received 13 November 2018; Received in revised form 17 August 2019; Accepted 1 September 2019
0019-8501/

Table 1
Key studies and knowledge gap.

Studies and data phases	BP	RP	UD	SD	Key findings and knowledge gap
1. Acquisition tools and techniques					
(Abbasi, Sarker, & Chiang, 2016)	√	×	√	√	The research at this point has a focus on data types and acquistion. The litreature sets the scene for big data frameworks (4Vs) and introduces the importance of strucutred and unstrucutred data, especially in relation to IOT devices, but primarily within the context of batch processing. There is hvery little discussion on the processing of data and no context is defined.
(Whitmore, Agarwal, & Da Xu, 2015)	×	×	√	×	
(Casado & Younas, 2014)	√	√	√	*	
(Kwon, Lee, & Shin, 2014)	×	×	*	√	
(Chen et al., 2012)	*	×	√	√	
(Uckelmann, Harrison, & Michahelles, 2011)	×	×	√	×	
2. Storage facilities					
(Watanabe & Nakamura, 2018)	*	×	√	×	This section has a focus on systems for scalability. Research investigates different infrastrucutre solutions (NOSQL, Hadoop, Apache) with a focus on technical hardware. No B2B or B2C marketing context. Plenty of wide reaching studies to inform theory.
(Hashem et al., 2015)	*	×	√	×	
(Whitmore et al., 2015)	×	×	√	√	
(Oliveira, Thomas, & Espadanal, 2014)	√	×	×	√	
(Grolinger et al., 2014)	√	×	×	√	
(Dean & Ghemawat, 2008)	√	×	√	√	
3. Analytical tools and techniques					
(Akhtar et al., 2019)	√	×	√	√	Real-time processing is now starting to be discussed as an analytical tool, in some quarters, however no context is given and discussion is very conceptual in scope. The major work in this area revolves around the use of Batch processing for decision making.
(Merla & Liang, 2018)	√	×	√	×	
(Wan et al., 2017)	*	*	√	√	
(Perez, Birke, & Chen, 2017)	√	×	×	√	
(Van Der Veen, Van Der Waaij, Lazovik, Wijbrandi, & Meijer, 2015)	×	√	×	×	
(Gandomi & Haider, 2015)	√	*	√	*	
(Hazen et al., 2014)	*	×	×	√	
4. Insights for data-driven actions					
(Yang et al., 2019)	√	×	×	√	Here we find a gap, current research in this field is focused primarily on Batch processing, as demonstrated by the table SD data is utilized in every paper to support decsion making for insight. No work on programmatic marketing, again very conceptual. Real-time processing as an approach towards data analysis is relatively new and has only recently come to the fore as an area of research.
(Kitchens et al., 2018)	√	×	√	√	
(Cao et al., 2018)	√	×	*	√	
(Arunachalam & Kumar, 2018)	√	×	×	√	
(Akhtar, Khan, Tarba, & Jayawickrama, 2017)	×	×	√	√	
(Chen et al., 2015)	√	×	*	√	

Batch processing (BP); Real-time processing (RP); Unstructured Data (SD); Structured Data (UD); × for none, √ for Yes, and * for "Yes, but not enough".

which we next briefly discuss. Firstly in identifying the practical gap, which influences decision making for B2B industrial marketing organizations, we highlight the variance between two key big data processing types; batch processing (Affetti, Tommasini, Margara, Cugola, & Della Valle, 2017; Casado & Younas, 2014; Grolinger et al., 2014) and real-time processing (Casado & Younas, 2014; Li et al., 2018; Wan et al., 2017). Secondly, working on the literature gap, our research in this paper has identified a significant body of research which discusses analytics (Cao et al., 2018; Järvinen & Karjaluoto, 2015; Nunan, Sibai, Schivinski, & Christodoulides, 2018; Yang, See-To, & Papagiannidis, 2019) underpinned by batch processing for decision making within industrial marketing organizations (see Table 1), this gap we argue suggests that there is very little discussion on real-time processing in a B2B industrial marketing context. Therefore, this paper intends to discuss the potential implementation of real-time processing within programmatic marketing to minimize cost, increase efficiencies and provide better targeted customer services for B2B industrial marketing organizations.

In supplementing the gap in the literature, this paper also problematizes literature sources that assist us in integrating our methodological approach, a mechanism of identifying research gaps and providing relative contributions (Nicholson, LaPlaca, Al-Abdin, Breese, & Khan, 2018; Sandberg & Alvesson, 2011). As the links between big data and programmatic marketing emerge, this strategy assists us in investigating the current knowledge gap and provides directions for future research. Specially, we challenge the field assumption of big data analytical batch processing that accepts past historical data as a mechanism to inform future actions, strategies and plans (Chen, Preston, & Swink, 2015; Kitchens, Dobolyi, Li, & Abbasi, 2018; Nunan et al., 2018; Yang et al., 2019). In some literature sources, for example supply chain management, the use of historical data is essential in developing clear forecasting approaches (Akhtar et al., 2019; Cao et al., 2018; Caputo et al., 2016). We argue that big data batch processing, which utilizes historical data to make decisions, places significant barriers in the implementation of programmatic marketing for B2B online display advertising, potentially creating a scenario where the data is obsolete before a decision has been made, we refer to this as *past loadedness*. For B2B industrial marketing managers, current practice in the purchasing of online display advertising relies heavily on google pay per click adverts, which are based on user search history (Erevelles, Fukawa, & Swayne, 2016), clickstreams (Erevelles et al., 2016) or social media likes or tweets (Yang, Lin, Carlson, & Ross, 2016). In moving away from these reactionary approaches to online display advertising, this paper proposes a challenge to the assumption of *past loadedness* with the counter-proposition that real-time data can significantly improve accuracy, efficiency and speed for *future loaded* online B2B programmatic marketing decision making.

In developing the paper envisioning as an approach was utilized to recognize and identify something that has yet to be established (MacInnis, 2011). Within this approach MacInnis (2011) argues that the goal is to introduce a new "*construct, theory, procedure, domain, discipline, or aspect of science that has yet to be apprehended or given serious study*", hence the chosen approach of envisioning to investigate big data processing and B2B programmatic marketing is consistent with theory blending.

We therefore pursue the following contributions. *First,* we identify a neglected gap in the literature from which we propose a real-time framework for B2B programmatic marketing. *Second,* through a problematization approach, we highlight the challenges of decisions made with historical datasets and bring forward the notion of "*future loadedness",* which we define as decisions that are made in real-time for effective B2B targeting. *Third,* by undertaking an interdisciplinary approach we bridge certain assumptions between two close but distinct areas of literature (i.e. big data and programmatic marketing), and highlight the challenges and issues of data processing for industrial marketing. In meeting these contributions and identifying the knowledge gap, we identify the following two relative research questions: *1. What are appropriate big data processing approaches for B2B industrial marketing for real-time display advertising; 2. How does structured and unstructured data influence big data processing techniques?* In meeting these two research questions this paper provides a potential framework for the development of real-time processing within programmatic marketing, for industrial marketing managers.

In structuring this paper, we take a non-formulaic approach, freeing the researcher from standard patterns in their work which constrain imagination and creativity (Alvesson & Gabriel, 2013). Within this non-formulaic approach Alvesson and Gabriel (2013) advocate the use of polymorphic research to "guide the practices of authors, reviewers and editors", this approach encourages authors to develop dialogue across multiple disciplines and research subsets. In moving this paper forward we first outline our key literature discussion, we then develop our theory blending approach across the fields of big data processing and programmatic marketing, we also discuss the emerging issues of structured and unstructured datasets. In keeping with the spirit of a theory blending approach we then proceed to discuss big data principles within the context of programmatic marketing and explore real-time processing for the purchasing of online display advertising. Finally, in our conclusion, we revisit our research questions and discuss in detail our contributions, the limitations and future research directions.

2. Strategies to conduct this study

In order to challenge existing assumptions in the field of big data and programmatic marketing we utilize the approach knows as problematization (Nicholson et al., 2018); an innovative approach which through the development of research questions challenges existing theories and concepts within existing literature. Indeed, in the view of Sandberg & Alvesson (2011, p. 32), this approach is to try and "*disrupt the reproduction and continuation of an institutionalized line of reasoning"*. In addition to this researchers (Nicholson et al., 2018) describe problematization as a deliberate calculated approach to articulate existing assumptions and challenge them with unique or novel ideas. The approach, while relatively new, has gained a certain level of momentum in the academy with researchers such as Bell, Kothiyal, and Willmott (2017) using it to challenge and examine methodologies and their rigor in globalized management research, focusing on the debates about the globalization of management research, while Dyer (2017) has used it to investigate strategies towards megaproject success. Other notable examples include Nicholson et al. (2018) who carry out a detailed analysis on the framing of contribution by highlighting different approaches taken by researchers in providing this framing.

This paper employs gap-spotting which is a subset of neglect spotting, and the strategy of problematization also known as assumption challenging (Nicholson et al., 2018). The motive for employing these specific strategies is to firstly challenge current thinking in this area and secondly develop revelatory contributions which can create new ways of knowing. This approach does not have to be revolutionary or ground breaking, in many cases it can challenge moderate assumptions within existing theories or intellectual traditions (Sandberg & Alvesson, 2011). Therefore, in Fig. 1 we highlight the use of gap spotting and assumption challenging in developing our contributions in the areas of big data and programmatic marketing for industrial B2B organizations:

In order to undertake a problematization and gap spotting approach the authors investigated the three main databases in this area (IEEE, Scopus, Science direct) and scraped the google scholar search engine. For accuracy and timeliness articles which had been published in the last 10 years within the areas *of Acquisition tools and techniques; storage facilities; Analytical tools and techniques; Insights for data driven action* were used. This is a framework based on the work of Casado and Younas (2014) who developed it for big data analysis and processing. Papers with a focus on big data, batch processing, real-time processing and programmatic marketing were selected (Akhtar et al., 2019; Hashem et al., 2015). The two strategies within both the review of the three databases and the google scholar search engine provide insight into the current literature landscape. Therefore, at the nexus of big data and programmatic marketing, this paper argues that there is a lack of research in the area of B2B industrial marketing. We also challenge the current literature in this area and argue that current research unquestionably accepts past historical data to inform future actions, strategies and plans. Therefore Table 1 below highlights our gap and our knowledge contribution in this paper:

Table 1 supports our gap-spotting and problematization approach; large amounts of data are collected, analyzed and stored but there is a gap in the literature in relation to real-time processing with unstructured data for B2B industrial marketing. Table 1 acts as a nexus for this paper highlighting the importance of data sources and big data processing and the lack of literature in the field.

3. Big data concepts

This interdisciplinary paper agrees with researchers who are of the view that within the field of analytics and big data there is a degree of saturation with the majority of the "low hanging" fruit having been claimed (Kitchens et al., 2018). The 'low hanging' fruit in this case is the notion that decisions are traditionally made in individual silos. Kitchens et al. (2018), argue that future research must take a more holistic approach by incorporating multiple structured and unstructured data sources to develop competitive advantage and decision making. Big Data can provide value based on the types of data collected, the larger the amount of volume the higher the chance of insight (Hashem et al., 2015). In the traditional sense researchers view big data through the prism of the volume of data collected. There are additional characteristics that build on this, for example in the view of Hashem et al. (2015), there are 4Vs of big data which they describe as *volume, variety, velocity and value.* While other researchers (Akhtar et al., 2017; Gandomi & Haider, 2015; Kunz et al., 2017) define various derivatives of these characteristics, which are generally regarded as the four common elements essential in the definition, maintenance, development and analysis of big data. Volume refers to the amount of data generated and the mass quantities in which it is collected. This initial aspect is critical and is, in many cases, closely associated with the notion of big data (Erevelles et al., 2016). Variety in this context refers to the different types of data collected and the various sources utilized. So for example Internet of things (IOT) in this context is crucial as it provides data from a vast variety of interconnected devices (Akhtar et al., 2017). Velocity is used to describe the speed at which the data is collected, this speed has a direct influence on how quickly the data can be analyzed to develop a strategy (Ahmed et al., 2017). The final V is value, which concerns insight and the process of discovery, where the focus is on obtaining hidden value. Others believe (Akhtar et al., 2019; Kitchens et al., 2018) that value comes when big data improves business performance.

Our review of the literature (Table 1) suggests that there is subtle shift in how organizations are making decisions, there is movement away from ad-hoc silos within the area of marketing towards the use of big data (Abbasi et al., 2016), IOT (Falkenreck & Wagner, 2017), social networking (Clemons, Dewan, Kauffman, & Weber, 2017; Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017; Yang et al., 2016) and analytics (Akhtar et al., 2017; Chen et al., 2012; Zhou et al., 2018). Thus, we argue that this subtle shift indicates an evolution in marketing research and practice for B2B organizations. Hence, in this context, underpinned by our theory building (Okhuysen & Bonardi, 2011), we suggest that the role of big data is becoming a significant disruptor in online and offline marketing approaches (Calder, Malthouse, & Maslowska, 2016). Some examples of disruption and the use of data in marketing include Falkenreck and Wagner (2017), who have developed research which investigates the use of the IOT to enhance buyer-seller interactions. Cao et al. (2018) explore the usage of analytics to make

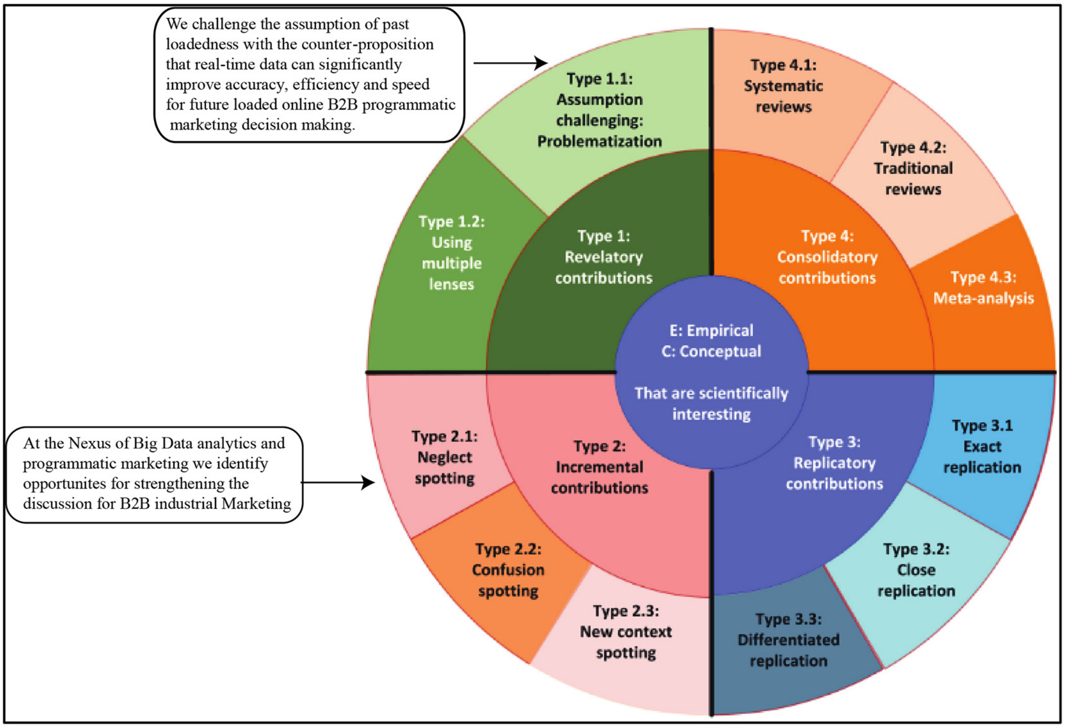

Fig. 1. Neglect spotting and Assumption challenging (Nicholson et al., 2018).

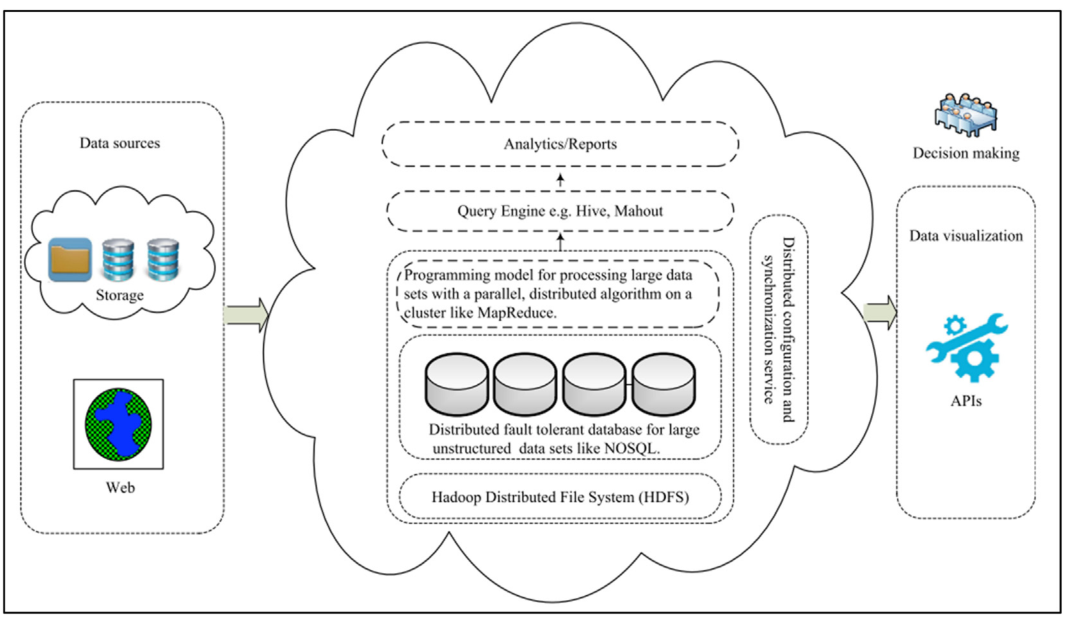

Fig. 2. Hashem et al. (2015) Cloud computing usage in big data environments.

strategic decisions, while Xu et al. (2016) discuss big data analytics for new product development. In addition to this there is other research which to a lesser extent discusses analytics within the context of real-time online decision making (Stevenson, 2015; Yuan, Wang, Li, & Qin, 2014).

This disruption shows signs of accelerating, as devices, audiences and organizations continue to become integrated data points, the implementation of solutions which converge these metrics will become prominent. In some technical scenarios this acceleration is taking place without human intervention, for example artificial intelligence and cognitive computing are currently being investigated and prototyped as potential tools for online marketing decision making (André et al., 2017; Chen, Argentis, & Weber, 2016). Hence, this evolution and disruption offers significant opportunities for B2B online display advertising, especially in relation to analytical tools and data-driven insight. However, any disruption and opportunity for B2B online display advertising needs to take into consideration the platform of architecture (Oliveira et al., 2014) and the key data sources for analysis (Gandomi & Haider, 2015). We now develop our discussion around infrastructure, and big data processing techniques.

3.1. Cloud based systems

This paper argues that the advent of cloud-based technology is one

Table 2
Structured and unstructured data benefits for B2B marketing.

	Structured data	Unstructured Data	Benefits for B2B programmatic marketing
Characteristics	Pre-defined data models Primarily text based Easy to search Easy to update	No pre-defined data model Text, images, sound, video, or any format Difficult to search & store	No model definition allows for a wider range of data to be collected and used in asset creation and decision making
Storage	Relational Databases Data Warehouses	NoSQL databases Data Warehouses Data Lakes	Complimentary data collection methods allow for data to be collected from multiple data sources.
Capture	Humans Machines	Humans Machines IOT	The use of IOT, humans and online machines provides rich, varied datasets that provide additional understanding of user requirements
Use case applications	CRM, ERP, Inventory, Supply chain management	Social media networks, Video uploads, photo gallery, voice notes,	Different datasets allow for unique asset creation so no two users may ever see the same B2B message
Data examples	Dates, Phone numbers, Credit card details, Address	Social Media, Keyword searches, clickstreams, YouTube videos	Real-time data allows B2B marketing companies to create real-time advertisements which are relevant at the point of need
Relevant literature	(Casado & Younas, 2014; Gandomi & Haider, 2015; Yang et al., 2019)	(Akhtar et al., 2017; Casado & Younas, 2014; Gandomi & Haider, 2015; Uckelmann et al., 2011; Yang et al., 2019)	

of the most significant shifts in modern information systems architecture for both service and enterprise applications (Hashem et al., 2015; Watanabe & Nakamura, 2018). From purely an infrastructure perspective, cloud based computing can provide the raw processing power, scalability and visualization processes to create environments in which big data analytics and large datasets (Hashem et al., 2015) can be stored and utilized to support real-time organizational decisions. Hashem et al. (2015) view visualization as a process of resource sharing and isolation of underlying hardware to increase computer resource utilization, efficiency, and scalability. Integrated data silos based upon centralized infrastructure as proposed by Kitchens et al. (2018), are supported through cloud based systems, which provide the backbone for the input of multiple offline and online data sources, but also for organizational outputs in the form of dynamic decision making, and data visualization (Fig. 2).

Hashem et al. (2015) highlight cloud-based infrastructure as a driver for the use of big data alongside analytical capability powered and stored in the cloud. The internal framework provided by Hashem et al. (2015) is heavily dependent on the storage of data for processing within a batch environment, and hence from a B2B industrial marketing approach unsuitable for programmatic marketing. In addition, the work of Hashem et al. (2015) highlights the role of multiple online and offline data stores, which crucially are based on historical big data warehouses and data lakes methods (Liang & Liu, 2018). Cloud based computing has been utilized in several recent studies, for example Song et al. (2019) investigate the robustness of the social internet of things for customization manufacturing through the potential implementation of cyber and physical elements in all aspects of business processes.

Movement towards cloud-based infrastructure provides a platform for the development of scalable and real-time analytics (Hazen et al., 2014), this is contrary to the traditional centralized IT infrastructure model which is so prevalent (Kitchens et al., 2018). The cloud also provides advantages in terms of cost, skills and hardware benefits; internal IT infrastructure is expensive and without the necessary skills and budget can quickly become outdated if not regularly maintained, which is also the case with IT strategy, policies and procedures (Kitchens et al., 2018). The movement towards cloud-based visualization removes prohibitive factors around costs and scalability, creating new opportunities for innovative growth and efficiencies within B2B online display advertising purchasing (Hashem et al., 2015).

3.2. Structured and unstructured big data sources

Big data has four characteristics and is underpinned by two distinct data categories; the first is structured data, which is analytics ready and is only a small subset of a big data source, and the second category is unstructured data, which is the largest component, constituting 95% (Gandomi & Haider, 2015) of a typical big dataset. These two data categories have diverse characteristics, which require unique storage topologies, contrasting processing techniques and in many cases distinct visualization approaches. Classifying the data into these two specific categories has been necessary due to the various data acquisition strategies employed by organizations. For example, the amount of data collected from IOT devices (Balaji & Roy, 2016), social media applications, website metrics (Erevelles et al., 2016) and search engines has required many B2B companies to revisit their data storage approaches. These types of online data sources have precipitated a significant shift away from traditional data collection approaches for business operations (Kitchens et al., 2018), creating a scenario where marketing companies need to evaluate the usefulness of multiple data sources as part of their decision making process (Erevelles et al., 2016). Access to different types of data which measure different metrics are, according to Abbasi et al. (2016), critical for developing insight and value while encouraging organizations to adopt contemporary requirements for online ad-slot purchasing.

Within this rapidly changing environment IOT is an emerging unstructured big data point representing internet enabled devices, which can be equipped with sensors and network transmitting features which allow them to communicate and relay data to a centralized system (Akhtar et al., 2017; Whitmore et al., 2015). While some early researchers (Xia, Yang, Wang, & Vinel, 2012) take the view that these are devices with inherent 'ubiquitous intelligence', later research (Akhtar et al., 2017) in this field takes a slightly different view and argues that they are part of an interconnected infrastructure, not standalone intelligent decision making agents. IOT devices are advantageous for in context data collection and in many scenarios play a vital role in collecting data which is not readily available from traditional data sources, creating a seamless environment that blends the virtual and physical data points (Uckelmann et al., 2011). As a data point, IOT can collect information on transportation, temperature, transactions and GPS among others (Akhtar et al., 2019).

Table 2 defines the characteristics of structured and unstructured data and highlights the key components of size, capture and characteristics:

Unstructured data provides consumer insight which previously was not readily available via structured datasets, the utilization of social media behavior and user clickstream data alongside keyword searches in search engines are resources which provide insight into consumer thinking in real-time, while the user web session is active. This type of data cannot be stored or captured within structured data processes due to data structures sitting outside pre-defined data models (Casado & Younas, 2014). For example, in a typical online session a consumer types in a keyword with the expectation that the results will be relevant and timely within the context of their session (Lewandowski, 2008). Through the collection of key word data and or clickstream information, organizations can bid on the purchasing of popular keywords in prime locations (Lewandowski, 2008). These prices are based on algorithms which utilize historical data; in competitive environments, this data could be obsolete before a decision is made. This type of online data is viewed as non-structured and hence does not fit into traditional database systems, creating challenges for storing such data types alongside structured information such as scanner data, internal records, files and consumer buying behavior (Erevelles et al., 2016). These challenges are compounded by the need to process and analyze these datasets. Investigating issues around data processing and their role in marketing research, we identify two different categories, which are batch and real-time processing (Casado & Younas, 2014). For the purposes of this paper, we discuss both these elements as potentially suitable for dynamic decision making in real-time. We now discuss these in turn.

4. Big data processing

As the volume of big data continues to grow, and its usage as a tool for analytic insight continues to support organizations to innovate (Casado & Younas, 2014), the requirement for new methods, tools and techniques to derive value from the data is becoming ever more crucial. Extant research (Casado & Younas, 2014; Gandomi & Haider, 2015) in this field has covered different processing paradigms, for example Hashem et al. (2015) discuss batch processing in cloud based systems using MapReduce, NOSQL and HDFS. In addition to this Wan et al. (2017) discuss the use of HDFS to receive uploaded historical data. The literature investigated as part of this paper covers a multitude of areas and context, and as Table 1 shows there is very little literature on real-time processing as a paradigm for B2B online display advertising decision making. Hence, the next stage in this paper is to investigate the two most popular approaches towards big data processing; batch and real-time processing.

4.1. Batch processing

The first popular method for processing big data is known as batch processing, this is typically used in applications where data naturally fits in a specific time window (Van Der Veen et al., 2015). As a mechanism it is defined as a highly efficient method for processing large volumes of data which has been collected over a significant period of time (Casado & Younas, 2014). The continued collection and storage of this type of data creates datasets which are extremely large, expensive to store and require significant infrastructure to manage and to derive value (Oliveira et al., 2014). Batch processing from a computer science perspective is designed to support the collection and storage of structured and unstructured datasets, creating a certain level of flexibility for future strategy discussions (Gandomi & Haider, 2015). Data captured from social media, click behavior, search engine searches, locations and smart devices (Hashem et al., 2015) can be stored alongside internal data such as transactions and sales information. However, the issue of data integrity and timeliness is always a concern and in scenarios where such large amounts of data are collected the majority of infrastructure is developed around a cloud based system to ensure active preventive maintenance (Wan et al., 2017). The storage of structured and unstructured data formats can cause schematic challenges, making the processing of data expensive and difficult, hence due to these challenges researchers describe big data as noisy, difficult to integrate, unclean and in its current state, of little strategic value (Casado & Younas, 2014; Kitchens et al., 2018).

To meet these challenges and to consider the role of batch processing, there are software, frameworks and concepts that have been specifically designed to analyze big data, the most popular being Hadoop. Hadoop is a layered structure to process and store massive amounts of data, a solution created and released by the Apache Software Foundation (Hashem et al., 2015), which utilizes the Hadoop Distributed File System (HDFS) for data processing and internal storage. Hadoop from a conceptual perspective highlights the difference between big data stores (NOSQL) and the traditional relational databases software (RDBMS). The traditional Relational database (RDBMS) cannot handle unstructured data sources, in scenarios such as this a non-relational database such as Hadoop is used (Shrivastava & Shrivastava, 2017). There are examples where Hadoop has been utilized as a solution for batch processing. Shrivastava and Shrivastava (2017) utilize Hadoop as a batch processing tool to derive insight from large datasets such as historical weblogs, past transactional data and sales data to develop customer profiles for traditional marketing activities.

Hadoop is the core component in the data analysis process and, through the MapReduce framework, can support the processing of large datasets using distributed algorithms (Merla & Liang, 2018). In a MapReduce environment there are two key functions; Map () and Reduce (), the Map function acts like a filter, grouping and sorting data, also referred to as data cleansing, whilst the Reduce function aggregates, and visualizes the data through a summary for the performance of an action (Dean & Ghemawat, 2008; Merla & Liang, 2018). Hence the role of these two functions within batch processing is to ensure the scalability of big data. However, in achieving this scalability there is an explicit trade off in accuracy for latency (Perez et al., 2017). This latency is due to the dependency on the MapReduce functions alongside a myriad of data inputs. The additional complexity and size of the data blocks in a Hadoop framework creates delays in processing and output, impacting on accuracy and speed (Perez et al., 2017). In order to visualize big data batch processing and some of the challenges in Fig. 3 we highlight the process of big data input, storage, processing and output:

While Fig. 3 does not link with B2B programmatic marketing, it outlines the overall batch process in its current form, our view for industrial marketing is the need to supplement this approach with data for real-time decision making. Also important to note is that for successful batch processing and the analysis and scalability of large datasets, MapReduce is a critical component (Dean & Ghemawat, 2008). Based on a lisp primitive, the map and reduce function works where users specify a map function to each logical “record”, the next stage is the computation of an intermediate key/value pairs (filter/combine), which then apply a reduce operation to all values that share the same key (Dean & Ghemawat, 2008; Shrivastava & Shrivastava, 2017). The map and reduce function have significant optimization and processing capabilities to develop efficiency in the analysis of large data providing long term value and insight on historical data. Thus, it is clear that the MapReduce paradigm, while effective in a batch processing framework, creates latency during the data processing phase. Other issues and challenges focus on MapReduce's fault tolerance implementation, which writes the results of the Map phase to local files before sending them to the reducers, as highlighted in Fig. 3. This causes multiple complications not least the significant high-overhead file systems (HDFS) which continually read-write from local files creating significant latency to the processing pipelines (Grolinger et al., 2014).

Hadoop and MapReduce are currently utilized within big data frameworks as a proposed solution for the analysis of data in order to identify historical patterns and trends (Casado & Younas, 2014). As highlighted in Table 1, there is a significant amount of research already conducted in this area to investigate big data batch processing. This approach to online data decision making allows for the processing of terabytes of data on thousands of machines, to develop a truly holistic customer view (Dean & Ghemawat, 2008). These datasets have a significant influence on organizations and the decisions they make in

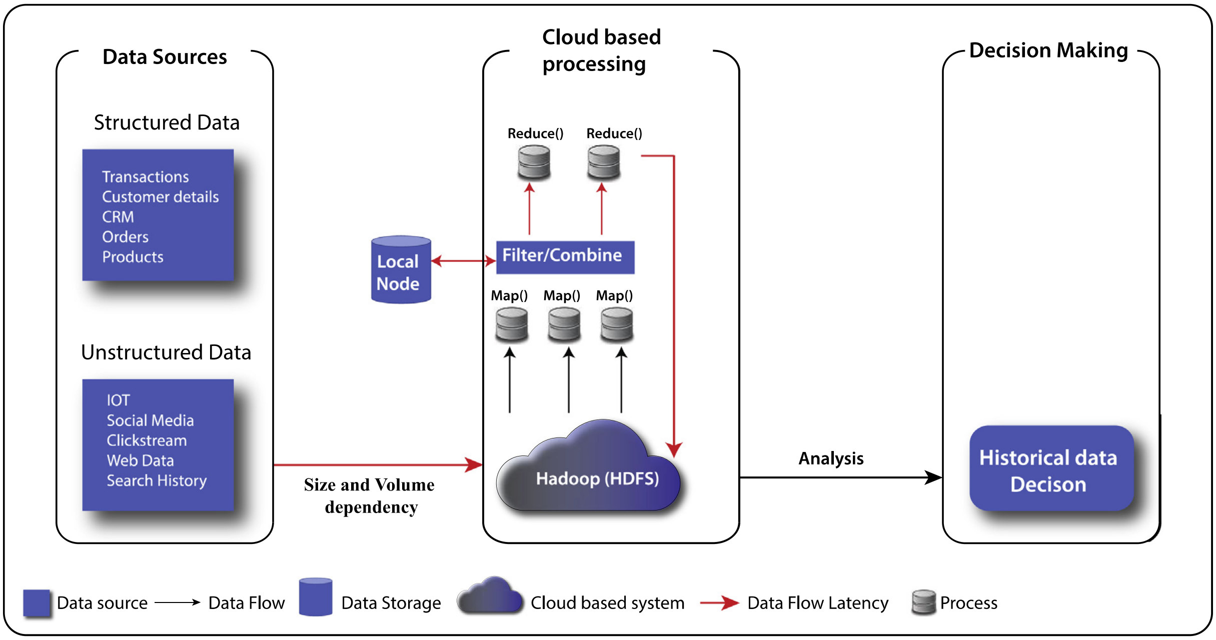

Fig. 3. Batch data processing for decision making.

relation to strategy, forecasting and consumer behavior etc. In challenging the literature at this stage, the use of batch processing for real-time B2B online display advertising decision making is unsustainable due to the focus on large historical datasets which are static and stored in local and remote storage causing latency in the decision-making process.

4.2. Real-time processing

In contrast to a Hadoop installation, there are other solutions that offer a different approach to the analysis of structured and unstructured datasets. Real-time processing is defined as an approach that requires a continuous stream of inputs for the processing and outputs of data (Casado & Younas, 2014). Within a real-time environment, organizations need to consider an approach that is fast, efficient and can handle multiple data types and formats; in this scenario data is seen as dynamic, constantly changing and updating based on the current environmental factors. The objective is to analyze and process in a small (or near real) period with minimum latency (Casado & Younas, 2014). It is at this point that the difference between batch processing and real-time processing becomes relevant. Batch processing is designed for rigorous results from big datasets, which have been collected over a period of time, while real-time processing is designed to accept continuously updated datasets for instantaneous decision making, within the latter aspect speed and efficiency are central to success (Casado & Younas, 2014; Kitchens et al., 2018). Akhtar et al. (2017) also discuss the importance of dynamic data in supporting organizational decision making for operational agility. They define dynamic data as the ability of the organization to cope with demands and changes both internally and externally (Akhtar et al., 2017). Arunachalam and Kumar (2018) support this view and argue that different data clustering approaches in the UK hospitality industry provide data driven insight for customer segmentation and target markets. In both these examples batch processing is central to the analysis of the data for decision making.

There are a wide range of processing platforms available which lend themselves to real-time processing. For this paper we investigate the use of Apache Storm, a solution developed as an open source initiative by the Apache foundation, it has a unique and robust reputation in the industry and a very active developer community. The open source nature of Storm has created a very flexible solution, allowing for additional code on demand and creating opportunities for developers to program their own real-time solutions if required (Van Der Veen et al., 2015). The movement towards real-time processing is being recognized in some quarters as an effective analytical tool. For example Wan et al. (2017) acknowledge this paradigm as an effective tool in active preventative maintenance within manufacturing, while Erevelles et al. (2016) discuss the concept of real-time processing in terms of data velocity within the context of big data. The research of the paper in this area expands this work, while these examples touch on real-time processing, it is still used in those contexts as a preventative tool or stored in a big data set on a local file store for batch processing at a later date. According to our research in this area, and the work we have identified in Table 1, there is a significant gap in discussing real-time processing for the development of decision making based on structured and unstructured data. Our technical illustration in Fig. 4 highlights the process of real-time processing within the solution of Apache storm, and where relevant the diagram highlights issue of latency, dependency and data flows.

In Fig. 4 the technical illustration for real-time data processing approach is underpinned by Apache storm, within a cloud based ecosystem. We also link this to B2B programmatic marketing to highlight the difference between historical and real-time output; as shown above, a real-time output leads to B2B real-time decision making. In addition, this infrastructure has a focus on speed and efficiency to minimize latency and delays in decision making, hence additional storm workers in this type of scenario would speed up the data decision making process. The nimbus (master node) controls the work flow and allocates tasks to the supervisors (worker node) based on need, the more supervisors a nimbus has access to, the quicker data can be processed. This workflow is influenced by the amount of data that streams into apache storm, here there is a potential for latency, the more real-time streams inputting data, the higher the risk of delay, as the balance between having enough streams to make real-time decisions and avoiding latency is very fine. In an Apache storm framework data is processed in mini-batches and is not directly stored in a database as in normal batch systems, this approach allows opportunity for direct real-time decision making or storage in a NoSQL database (HDFS), which can cause latency, this we illustrate above (Watanabe & Nakamura, 2018). The key to this approach is the early identification of B2B structured and unstructured data sources, this is done via the real-time constant streams.

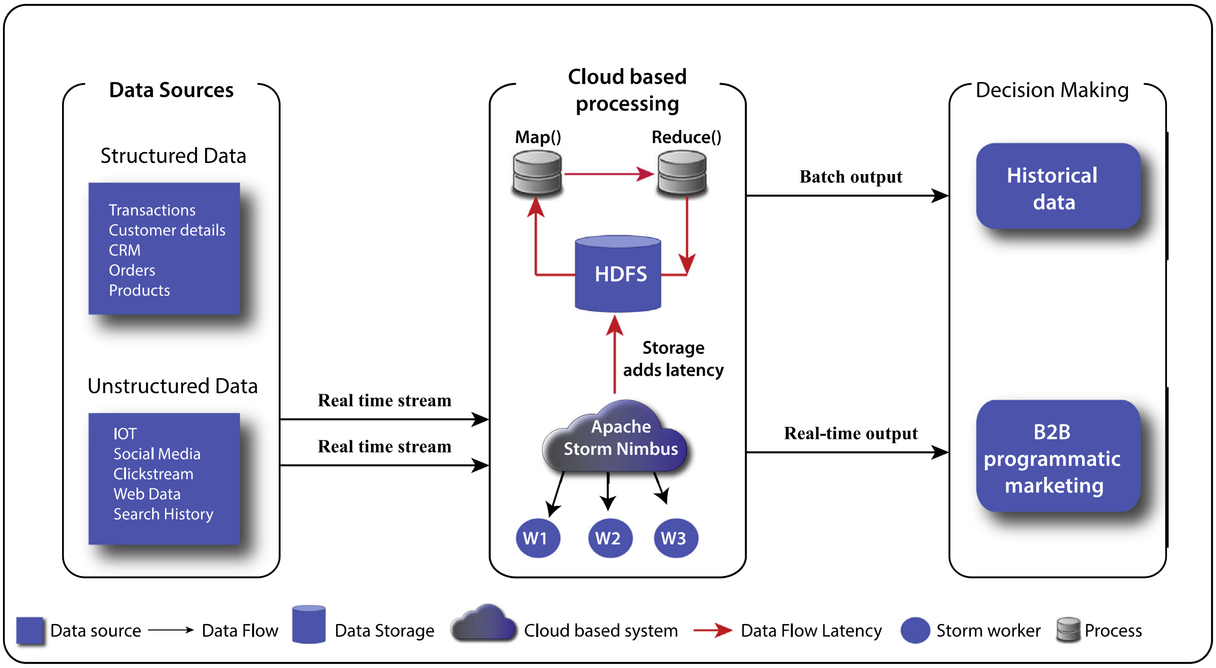

Fig. 4. Real-time data processing.

In addition, the divergence of data collection mechanisms creates opportunity for the identification of additional data value incorporated from data sources (unstructured), which is not the norm for many organizations (Gandomi & Haider, 2015). Thus, the real-time processing approach allows for the creation of new services in different areas. In this scenario we discuss the potential for real-time approaches for B2B online display advertising purchasing.

5. B2B programmatic marketing and relative implications

In Table 1 we identify our key knowledge gap in this paper, and contend that there is very little research within B2B industrial marketing which explores the role of programmatic marketing for targeted, real-time online display advertising. In response to this lack of research our conceptual paper argues for an infrastructure which supports real-time processing frameworks and can make decisions on the 'fly' based on user web sessions and preferences. Extant research in this field has explored online digital marketing strategy for the development of contextualized, targeted and highly unique content for its B2B customers (Jobber & Ellis-Chadwick, 2012; Mani & Chouk, 2016). However, there has been very little or no work on the impact of programmatic marketing in the online display process.

In this paper we bring a unique perspective to future B2B marketing research. Our work on the investigation of batch and real-time processing has allowed us to identify the Apache storm framework for a variety of B2B industrial marketing solutions, namely programmatic marketing for online display advertising. Most other approaches such as marketing analytics and predictive marketing are better equipped for batch processing (Kitchens et al., 2018), based on historical data for forecasting purposes, which while valuable occur after the user has exited the current session, and thus could result in the loss of revenue. Real-time processing in the context of display advertising purchasing within programmatic marketing is new and a very niche approach to industrial marketing, making use of data feeds in streams passing constantly through the Apache storm framework for B2B marketing decisions. Within this approach user needs are identified in real-time and structured and unstructured data streams are used to formulate marketing decisions about individual users or B2B organizations. Any latency at this point can cause delays in the response process and customer attention within a web session could be lost. Hence, dealing with data and decision making in such a manner requires a process that is automated for organizations and highly targeted using new location based techniques (Uckelmann et al., 2011) and new marketing approaches in addition to traditional approaches such as CRM onboarding, IP targeting and geo-targeting. These approaches are complex and require the compression of complex decisions into micro seconds, as articulated in this paper, this requires the use of big data with a scalable infrastructure such as a cloud-based environment (Hashem et al., 2015).

The conceptual approach in this paper is the natural evolution of online marketing. As the marketing process becomes competitive, website buying, ad-slot buying, online publishing, customer profiling, targeting, search engine optimization and content generation to name a few, will go through a process of significant disruption. The marrying of real-time processing within programmatic marketing provides an environment of targeted advertising with clear transparency. This overlay of the two distinct approaches can only be fully measured if the data is generated and benchmarked at volume, and the influence on consumer behavior is measured at the point of purchasing. This approach of distinct targeting moves away from current online marketing approaches. For example, the use of ad impression placement, pay per click purchasing and online banner buying are illustrations of tactics that are unpredictable, as predicting visits to the publishers landing page is an imperfect science (Balseiro, Feldman, Mirrokni, & Muthukrishnan, 2014). This imperfect science suggests that too often B2B marketing companies take a predictive analytics approach to batch processing with narrow datasets. This creates challenges for targeted, relevant and timely decision in a manner that can resonate with customers when they need them in the current real-time session. A technical illustration of B2B online display advertising within a programmatic marketing environment is highlighted below:

Fig. 5 illustrates the notion of real-time data to support B2B programmatic marketing decision making at the point of inspiration (consumer session). The vision and contribution of this paper towards dialogue argues that real-time processing implemented alongside a programmatic marketing framework can potentially create an opportunity to take a forward-looking approach in the serving of adverts in real-time. In Fig. 5 we highlight that data can be stored during the

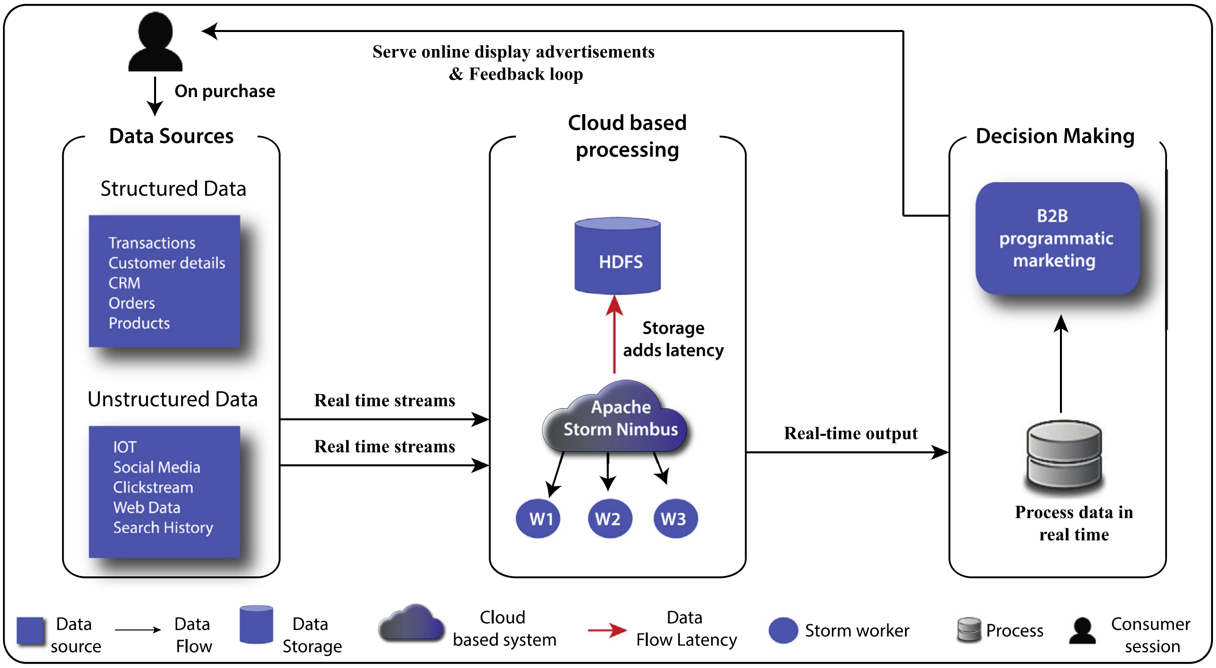

Fig. 5. Real-time processing for B2B Online display advertising.

processing stage but this extra function can add latency in the programmatic marketing process, the focus on output and speed is critical. The key aspect of this approach is also the integration of consumer feedback on the value of the data being collected, ultimately poor-quality data can be discarded and the feedback loop can then serve better adverts within the session.

From a practical implementation perspective a programmatic approach to online advertising sales is the optimal way of promoting products in the online environment (Gonzalvez-Cabañas & Mochón, 2016; Li et al., 2018; McGuigan, 2019), directly linked to speed and efficiency that help managers and decision makers to implement their contemporary tools, and use stream processing as an essential tool for real time data collection and decision making. Our proposed solution argues that within real-time online display advertising there is significant scope to minimize latency, increase data collection and minimize disruption in customer targeting. Gonzalvez-Cabañas and Mochón (2016) in their paper focus on online advertising purchasing within five-minute intervals to analyze data segments according to metrics such as speed of impressions per hour, speed in this scenario and the display of adverts is key based on platform and profile metrics. This approach differs from our perspective where the focus is on real time decision making and purchasing, a five-minute interval from the proposed solution would add delays to the decision-making process. The results of Gonzalvez-Cabañas and Mochón (2016) based on the use of a stream confirm the low-latency, through an increased in successful advert placements rather than real time data collection and instant decision making.

Research in this area also suggests that in practical terms while initial costs maybe expensive, over the long-term implementation of programmatic marketing will outperform traditional approaches to online decision making (Gonzalvez-Cabañas & Mochón, 2016; Li et al., 2018; McGuigan, 2019). For example, in the context of advertisement inventory allocation Li et al. (2018) formulate ad-inventory allocation over three optimized levels creating greater efficiency and a cost effective approach, proposing a targetted user centered approach. The proposed framework would provide a seamless fit for the three levels. Thus, both Li et al. (2018) and Gonzalvez-Cabañas and Mochón (2016) support the central premise of this paper, albeit, on a smaller scale that successful practical implementation is dependent on increased speeds and capacity that will drive commodification in generating, processing, and coordinating flows of information and commerce. This is supported by McGuigan (2019) who argues that programmatic marketing will speed and refine the automation, algorithmic decision-making, and general developments in ad-tech for the twin goals of selling audiences and selling sponsors' products, leaving behind traditional media approaches that in comparison are unprofitable and slow.

Thus, the movement towards a real-time processing framework underpinned by integrated data is critical in improving marketing decisions, more effective and cost efficient, with potentially less errors. This view is supported by Shrivastava & Shrivastava, (2017), who argue that big data is designed to improve decision making, reduce errors, increase efficiency, make faster policy decisions and provide innovative solutions. The proposed technical framework in Fig. 5 attempts to simulate an information flow for structured and unstructured data, for the process of decision making that can be undertaken quickly, underpinned by good quality data regulated by the feedback loop.

6. Conclusions and future research directions

6.1. Our contributions

Our broad aims for this paper were threefold. First, through gap spotting and problematization we provided a general big data processing framework for the development of B2B programmatic marketing. In the second we sought to identify, expand, explore and understand the implications of structured and unstructured data within big data processing, and we propose the term "*future loadnesses*". In the final contribution we aimed to provide interdisciplinary literature where we bridge the gap between the fields of big data and programmatic marketing, and highlight the challenges of processing for industrial marketing. These contributions have created a unique opportunity to stimulate dialogue and enhance our understanding in the potential use of programmatic marketing within industrial marketing. Our problematization and gap-spotting approach has done much to highlight the key differences between historical and real-time data, and points the way forward in future B2B marketing actions. While authors such as Li et al. (2018) have started to investigate the use of programmatic marketing, much of their work still has a focus on batch data which is historical in nature. As argued in this paper, this type of decision-making approach fits the need in specific scenarios, but is limited when decisions need to be made in the current user session. Therefore, creating challenges for consumer targeting and purchasing causing additional expense and poor-quality targeting.

6.2. The research questions

Earlier in the paper we identified the following two research questions to guide and develop this paper: 1. What are appropriate big data processing approaches for B2B industrial marketing for real-time display advertising?; 2. How does structured and unstructured data influence big data processing techniques? For the first research question, we investigated batch and real-time processing as approaches for real-time display advertising, while also exploring the latency challenges and issues around data storage. This research is of the opinion that while both batch and real-time processing are suitable for the storage and interrogation of big data, only real-time processing within an apache storm configuration is really suitable for B2B online display advertising within a programmatic marketing environment. The primary reason for this conclusion is the difference between how the two approaches interrogate and process data. Batch processing with an optimized Hadoop and Map-reduce solution is appropriate for the storage and manipulation of data for forecasting and long-term strategy purposes, and not really suitable for analysis of decisions that are required in real-time. The capturing of customer attention at the point of inspiration requires an Apache Storm configuration to provide relevant and timely adverts within the current user session.

For the second research question a significant amount of time was spent exploring big data sources, data types and the very nature of big data through the prism of the 4vs. Structured and unstructured datasets have provided organizations with a myriad of opportunities to capture information on locations, transactions, sentiments and web views, all for the prospect of creating unique marketing opportunities (Yang et al., 2016). While these new data points add significant value to organizations there are, however, challenges involved in cleaning and storing unstructured data to ensure it is usable alongside structured datasets for accurate decision making. The presence of noise in a dataset is a particular issue, which García-Gil, Luengo, García, and Herrera (2017) define as *"the partial or complete alteration of the information gathered for a data item, caused by an exogenous factor not related to the distribution that generates the data"*. Noise in a dataset will lead to excessively complex models impacting on performance with increased latency with minimum value. Hence, with noise having such a big influence on datasets the influence on data processing and techniques is profound, especially in relation to delays. Thus, the challenge for marketing organizations is not collecting or storing the data in a centralized location, but cleansing the data in a manner which aids organizational use and strategic decision making, while minimizing latency.

6.3. Limitations and future research directions

We suggest that while the framework proposed in this paper offers a significant step in understanding the role of data for targeted real-time programmatic marketing, an understanding of data integration across internal and external networks alongside latency is an agenda which needs to be pursued. This integration of data at all levels across the organization creates conditions for targeted and specific B2B display advertising in programmatic marketing. An integrated approach creates software and hardware compatibility across multiple systems (Song et al., 2019), allowing organizations to make bids on display advertising in real-time which are then served to the customer instantaneously. Other advantages of integrated datasets include the retrieval and combination of different data types to be merged by stakeholders across systems. Integration allows flexibility and control across a range of business processes, moving away from ad-hoc data silos which can often be prone to security, ownership, editing challenges (Kitchens et al., 2018).

We acknowledge that this paper has limitations, which are mainly related to the conceptual view this paper has taken and while we propose a detailed framework, this paper lacks an agenda to discuss the limitations of the decision-making process. In the context of this paper decision making is instantaneous and driven by the structured and unstructured datasets, hence data overload and poor-quality data collection techniques can, as discussed, lead to poor decision making, poor targeting and an increase in costs. This also causes issues with delays in the decision-making process, increasing latency times. Future work in the area of decision making needs to have a strategy which aims to discuss and investigate these limitations, our framework has provided the basic fundamentals in this context but future discussion needs to focus on the implications of different data types and how they influence the speed of programmatic marketing delivery.

In conclusion, in this paper we argue that B2B programmatic marketing within real-time decision making is a natural evolution of the discipline. As marketing becomes more competitive, instantaneous real-time decision making will have a huge influence on B2B competitive advantage, in this context the development of this paper takes an interdisciplinary approach bringing together literature and ideas from computer science and marketing. Our three contributions, which we laid out at the start of this paper, challenge the assumption within research that indicates past historical data should inform future actions, strategies and plans. While in some scenarios this is a valid mechanism, for B2B online display advertising this is not suitable in a real-time programmatic marketing context. The conceptual framework proposed in Fig. 5 is the natural evolution of marketing, driven by data and underpinned by analytics.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems, 17*(2), 32.

Affetti, L., Tommasini, R., Margara, A., Cugola, G., & Della Valle, E. (2017). Defining the execution semantics of stream processing engines. *Journal of Big Data, 4*(1), https://doi.org/10.1186/s40537-017-0072-9.

Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in internet of things. *Computer Networks, 129*, 459–471. https://doi.org/10.1016/j.comnet.2017.06.013.

Akhtar, P., Frynas, J. G., Mellahi, K., & Ullah, S. (2019). Big data-savvy teams' skills, big data-driven actions and business performance. *British Journal of Management, 30*(2), 252–271.

Akhtar, P., Khan, Z., Tarba, S., & Jayawickrama, U. (2017). The internet of things, dynamic data and information processing capabilities and operational agility. *Preprint*, 1–32. https://doi.org/10.1016/j.techfore.2017.04.023.Pervaiz.

Akhtar, P., Tse, Y. K., Khan, Z., & Rao-Nicholson, R. (2016). Data-driven and adaptive leadership contributing to sustainability: Global agri-food supply chains connected with emerging markets. *International Journal of Production Economics, 181*, 392–401.

Alvesson, M., & Gabriel, Y. (2013). Beyond formulaic research: In praise of greater diversity in organizational research and publications. *The Academy of Management Learning and Education, 12*(2), 245–263. https://doi.org/10.5465/amle.2012.0327.

André, Q., Carmon, Z., Wertenbroch, K., Crum, A., Frank, D., Goldstein, W., ... Yang, H. (2017). Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs and Solutions,* 28–37. https://doi.org/10.1007/s40547-017-0085-8.

Arunachalam, D., & Kumar, N. (2018). Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making. *Expert Systems with Applications, 111*, 11–34. https://doi.org/10.1016/j.eswa.2018.03.007.

Balaji, M. S., & Roy, S. K. (2016). Value co-creation with internet of things technology in the retail industry. *Journal of Marketing Management, 2025*(1–2), 1–25. https://doi.org/10.1080/0267257X.2016.1217914.

Balseiro, S. R., Feldman, J., Mirrokni, V., & Muthukrishnan, S. (2014). Yield optimization of display advertising with ad exchange. *Management Science, 60*(12), 2886–2907. https://doi.org/10.1287/mnsc.2014.2017.

Bell, E., Kothiyal, N., & Willmott, H. (2017). Methodology-as-technique and the meaning of rigour in globalized management research. *British Journal of Management, 28*(3), 534–550.

Calder, B. J., Malthouse, E. C., & Maslowska, E. (2016). Brand marketing, big data and social innovation as future research directions for engagement. *Journal of Marketing Management, 32*(5–6), 579–585. https://doi.org/10.1080/0267257X.2016.1144326.

Cao, G., Duan, Y., & El Banna, A. (2018). A dynamic capability view of marketing

analytics: Evidence from UK firms. *Industrial Marketing Management*. https://doi.org/10.1016/j.indmarman.2018.08.002 September 2017.
Caputo, A., Marzi, G., & Pellegrini, M. M. (2016). The internet of things in manufacturing innovation processes. *Business Process Management Journal, 22*(2), 383–402. https://doi.org/10.1108/BPMJ-05-2015-0072.
Casado, R., & Younas, M. (2014). Emerging trends and technologies in big data processing. *Concurrency and Computation: Practice and Experience, 27*(8), 2078–2091. https://doi.org/10.1002/cpe.3398.
Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems, 32*(4), 4–39. https://doi.org/10.1080/07421222.2015.1138364.
Chen, H., Storey, V. C., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188. https://doi.org/10.1145/2463676.2463712.
Chen, Y., Argentis, E., & Weber, G. (2016). IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clinical Therapeutics, 38*(4), 688–701. https://doi.org/10.1016/j.clinthera.2015.12.001.
Clemons, E. K., Dewan, R. M., Kauffman, R. J., & Weber, T. A. (2017). Understanding the information-based transformation of strategy and society. *Journal of Management Information Systems, 34*(2), 425–456. https://doi.org/10.1080/07421222.2017.1334474.
Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107–113.
van Dieijen, M., Borah, A., Tellis, G. J., & Franses, P. H. (2019). Big data analysis of volatility spillovers of brands across social media and stock markets. *Industrial Marketing Management*, 1–20. https://doi.org/10.1016/j.indmarman.2018.12.006 December 2018.
Dyer, R. (2017). Cultural sense-making integration into risk mitigation strategies towards megaproject success. *International Journal of Project Management, 35*(7), 1338–1349.
Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001.
Falkenreck, C., & Wagner, R. (2017). The internet of things – Chance and challenge in industrial business relationships. *Industrial Marketing Management, 66*(August), 181–195. https://doi.org/10.1016/j.indmarman.2017.08.007.
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.
García-Gil, D., Luengo, J., García, S., & Herrera, F. (2017). Enabling smart data: Noise filtering in big data classification. *Information Sciences, 479*(2019), 135–152. https://doi.org/10.1007/s10815-006-9022-8.
Gonzalvez-Cabañas, J. C., & Mochón, F. (2016). Operating an advertising programmatic buying platform: A case study. *International Journal of Interactive Multimedia and Artificial Intelligence, 3*(6), 6. https://doi.org/10.9781/ijimai.2016.361.
Grolinger, K., Hayes, M., Higashino, W. A., L'Heureux, A., Allison, D. S., & Capretz, M. A. M. (2014). *Challenges for MapReduce in big data.* 182–189. https://doi.org/10.1109/services.2014.41.
Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems, 26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003.
Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98–115. https://doi.org/10.1016/j.is.2014.07.006.
Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics, 154*, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018.
Järvinen, J., & Karjaluoto, H. (2015). The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management, 50*, 117–127.
Jobber, D., & Ellis-Chadwick, F. (2012). *Principles and practice of marketing.* McGraw-Hill Higher Education.
Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Information Systems, 35*(2), 540–574. https://doi.org/10.1080/07421222.2018.1451957.
Kunz, W., Aksoy, L., Bart, Y., Heinonen, K., Kabadayi, S., Ordenes, F. V., ... Theodoulidis, B. (2017). Customer engagement in a big data world. *Journal of Services Marketing, 31*(2), 161–171.
Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management, 34*(3), 387–394. https://doi.org/10.1016/j.ijinfomgt.2014.02.002.
Lewandowski, D. (2008). Search engine user behaviour: How can users be guided to quality content? *Information Services & Use, 28*(3–4), 261–268.
Li, J., Ni, X., Yuan, Y., & Wang, F. Y. (2018). A hierarchical framework for ad inventory allocation in programmatic advertising markets. *Electronic Commerce Research and Applications, 31*(May), 40–51. https://doi.org/10.1016/j.elerap.2018.09.001.
Liang, T. P., & Liu, Y. H. (2018). Research landscape of business intelligence and big data analytics: A bibliometrics study. *Expert Systems with Applications, 111*(128), 2–10. https://doi.org/10.1016/j.eswa.2018.05.018.
MacInnis, D. J. (2011). A framework for conceptual contributions in marketing. *Ssrn, 75*(4), 136–154. https://doi.org/10.2139/ssrn.1845968.
Mani, Z., & Chouk, I. (2016). Drivers of consumers' resistance to smart products. *Journal of Marketing Management, 00*(00), 1–22. https://doi.org/10.1080/0267257X.2016.1245212.
McGuigan, L. (2019). Automating the audience commodity: The unacknowledged ancestry of programmatic advertising. *New Media & Society*. https://doi.org/10.1177/1461444819846449.
Merla, P. R., & Liang, Y. (2018). Data analysis using hadoop MapReduce environment. *Proceedings - 2017 IEEE international conference on big data, big data 2017, 2018-Janua* (pp. 4783–4785). . https://doi.org/10.1109/BigData.2017.8258541.
Mouzas, S., & Araujo, L. (2000). Implementing programmatic initiatives in manufacturer–retailer networks. *Industrial Marketing Management, 29*(4), 293–303. https://doi.org/10.1016/S0019-8501(00)00107-3.
Nicholson, J. D., LaPlaca, P., Al-Abdin, A., Breese, R., & Khan, Z. (2018). What do introduction sections tell us about the intent of scholarly work: A contribution on contributions. *Industrial Marketing Management*, 1–14. https://doi.org/10.1016/j.indmarman.2018.02.014 December 2016.
Nunan, D., Sibai, O., Schivinski, B., & Christodoulides, G. (2018). Reflections on "social media: Influencing customer satisfaction in B2B sales" and a research agenda. *Industrial Marketing Management, 75*(January), 31–36. https://doi.org/10.1016/j.indmarman.2018.03.009.
Okhuysen, G., & Bonardi, J.-P. (2011). The challenges of building theory by combining lenses. *Academy of Management Review, 36*(1), 6–11. https://doi.org/10.5465/AMR.2011.55662498.
Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information and Management, 51*(5), 497–510. https://doi.org/10.1016/j.im.2014.03.006.
Oswick, C., Fleming, P., & Hanlon, G. (2011). From borrowing to blending: Rethinking the processes of organizational theory building. *Academy of Management Review, 36*(2), 318–337. https://doi.org/10.5465/AMR.2011.59330932.
Perez, J. F., Birke, R., & Chen, L. Y. (2017). On the latency-accuracy tradeoff in approximate MapReduce jobs. *Proceedings - IEEE INFOCOM*. https://doi.org/10.1109/INFOCOM.2017.8057038.
Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: Gap-spotting or problematization? *Organization, 18*(1), 23–44. https://doi.org/10.1177/1350508410372151.
Shrivastava, G., & Shrivastava, S. (2017). Analysis of customer behavior in online retail marketplace using Hadoop. *International Journal of Innovative Research in Computer Science & Technology, 5*(5), 376–380. https://doi.org/10.21276/ijircst.2017.5.5.3.
Song, Z., Sun, Y., Wan, J., Huang, L., Xu, Y., & Hsu, C. H. (2019). Exploring robustness management of social internet of things for customization manufacturing. *Future Generation Computer Systems, 92*, 846–856. https://doi.org/10.1016/j.future.2017.10.030.
Stevenson, D. (2015). *Market-based audience costs in real-time bidding advertising systems*.
Uckelmann, D., Harrison, M., & Michahelles, F. (2011). An architectural approach towards the future internet of things. *Architecting the internet of things* (pp. 1–24). Springer.
Van Der Veen, J. S., Van Der Waaij, B., Lazovik, E., Wijbrandi, W., & Meijer, R. J. (2015). Dynamically scaling apache storm for the analysis of streaming data. *Proceedings - 2015 IEEE 1st international conference on big data computing service and applications, BigDataService 2015* (pp. 154–161). . https://doi.org/10.1109/BigDataService.2015.56.
Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., & Vasilakos, A. V. (2017). A manufacturing big data solution for active preventive maintenance. *IEEE Transactions on Industrial Informatics, 13*(4), 2039–2047. https://doi.org/10.1109/TII.2017.2670505.
Watanabe, S., & Nakamura, A. (2018). In A. Sieminski, A. Kozierkiewicz, M. Nunez, & Q. T. Ha (Eds.). *Integrated data access to heterogeneous data stores for IoT cloud BT - modern approaches for intelligent information and database systems* (pp. 423–433). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-76081-0_36.
Whitmore, A., Agarwal, A., & Da Xu, L. (2015). The internet of things: A survey of topics and trends. *Information Systems Frontiers, 17*(2), 261–274. https://doi.org/10.1007/s10796-014-9489-2.
Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems, 25*(9), 1101–1102. https://doi.org/10.1002/dac.2417.
Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research, 69*(5), 1562–1566. https://doi.org/10.1016/j.jbusres.2015.10.017.
Yang, S., Lin, S., Carlson, J. R., & Ross, W. T. (2016). Brand engagement on social media: Will firms' social media efforts influence search engine advertising effectiveness? *Journal of Marketing Management, 32*(5–6), 526–557. https://doi.org/10.1080/0267257X.2016.1143863.
Yang, Y., See-To, E., & Papagiannidis, S. (2019). You have not been archiving emails for no reason! Using big data analytics to cluster B2B interest in products and services and link clusters to financial performance. *Industrial Marketing Management,* (January), 1–14. https://doi.org/10.1016/j.indmarman.2019.01.016.
Yuan, Y., Wang, F., Li, J., & Qin, R. (2014). A survey on real time bidding advertising. *Service operations and logistics, and informatics (SOLI), 2014 IEEE international conference on* (pp. 418–423). IEEE.
Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., & Yan, X. (2018). Measuring customer agility from online reviews using big data text analytics. *Journal of Management Information Systems, 35*(2), 510–539. https://doi.org/10.1080/07421222.2018.1451956.

Abdul Jabbar is the Director of Learning Development at the University of Huddersfield. He has professional experience in a software development environment and has been working as an academic for the last 10 years. His areas of expertise encompass big data, business analytics, information systems, software development, blockchain and programmatic marketing.

Pervaiz Akhtar is Associate Dean of Graduate Studies and Professor of Management Systems, specialized in big data, business analytics, technology, mixed methods and modelling to optimize and de-risk operational performance, supply chains and business value. He is also an External/Visiting Professor in Big Data and Business Analytics/Management Science, IESEG School of Management (France). Due to his all-round excellence in leadership, research and teaching, he made to the list of 155 academics (out of 194,245 professors, 0.00080%) across all disciplines in the UK who earned their Professorship under the age of 35 as per HESA records (www.hesa.ac.uk) and became the youngest professor at multiple universities. He has also set consecutive promotion records. Capitalizing on over 15 years of academic and industrial experiences in public, private, and non-profit-making organizations (UNICEF, JCCP, Oxfam, MSF, Islamic Relief, Red Cross, high government officials, dairy industry, and FMCG companies such as Unilever and Reckitt Benckiser), his research has appeared in top-ranked journals such as *International Journal of Operations & Production Management, British Journal of Management, Business Ethics, International Journal of Production Economics, Supply Chain Management: An International Journal, Industrial Marketing Management (methodology), Transportation Research Part A: Policy and Practice, Journal of Business Research*, among others. He is the corresponding author and can be reached at pervaiz_khan972@hotmail.com or P.Akhtar@kent.ac.uk.

Samir Dani is Professor of Logistics and Supply Chain Management and Head of Department at the University of Huddersfield Business School. Professor Dani engages widely with industry and has research interests in supply chain risks, sustainability in supply chains, and the use of technology and business models. He is currently conducting research in the use of AI techniques and autonomous systems within supply chain decision making and logistics. He has published widely and has presented his research to both academic and practitioner audiences.