

Gizzle Bautista
CC53 MIS - S0261553
Email: gizzle.bautista@cquemail.com

COIT20253 Business Intelligence using Big Data
Assignment 1
Due Date: 22 April 2016

Lecturer: Meena Jha
Course Coordinator: Meena Jha

Table of Contents

Introduction	1
Big Data	1
Big Data and the Publishing Industry	1
Kobo Inc.	1
Online versus Offline Big Data	2
Selection Strategy for Big Data Application	3
Big Data Solutions Outcome	4
Technologies in Big Data Solutions	4
Business Impact of Big Data Solutions	5
Organisational Impact of Big Data Solutions	6
Conclusion	7
References	8

Introduction

With the uprising of Big Data solutions in different fields, the need to understand what it does and how it affects the way organisations do their businesses is vital for every IT professional. This report aims to provide its definition and examples of the publishing industry's use of Big Data. This report will also investigate the impact of Big Data solutions in the publishing industry. A positive and negative perspective on the solution will be discussed. Furthermore, Kobo Inc.'s use of Big Data will be presented as a case study sample. Relating all the literatures provided, this paper aims to provide a conclusive assessment on how Big Data solution is changing the publishing world through business intelligence analysis.

Big Data

In a white paper published by Kobo (2014), the company described Big Data as large data sets that when correct analysis is applied can be a tool for business forecasting. While Press (2014), contributor for Forbes magazine, provided several of his own definitions wherein he cited Big Data as the organisational attitude of combining data from different sources which could lead to better decision making strategies. Lastly, for MongoDB (2015), Big Data enables organisations to create new merchandises in order to be highly competitive and at the same time save money.

Big Data and the Publishing Industry

A decade ago, the only concern of book publishers is for its content to be bought while their sale status in physical stores are determined through bookscan (Davenport 2014). Nowadays, publishers, as content producers/packagegers/distributors, need personalised content delivery and content recommendations (Bright 2015). Wherein, analysing a single reader's journey and their use of multiple devices throughout is another spectrum to consider as per the author.

Kobo Inc.

Kobo Inc.'s journey in Big Data started in 2012 is originally started to solely help customers but later on made as a profit centre (Christensen 2013). The company website describe itself as "the world's fastest-growing eReading services" and DigitalPublishing101 (2015) differentiated Kobo from Amazon.com Inc. as having direct access to publishers and that it exclusively sells ebooks.

In the Big Data Innovation Summit 2013, Kobo's VP of Big Data presented how readers create massive amount of data. Christensen (2013) explained how readers' behaviour in store, such as visiting, searching, browsing and buying are example of data creation. Additionally, the readers' actions in their ebook readers produce data (Christensen 2013).

Online versus Offline Big Data

According to the MongoDB's (2015) published white paper, the decision in selecting which Big Data technology to use depends on how the organisation intends to use their data. If the organisation needs more real-time operational use cases then the company needs an online big data technology while a long-running offline analysis require offline big data solutions (MongoDB 2015).

Figure 1 below shows MongoDB's (2015) comparison of Online and Offline Big Data:

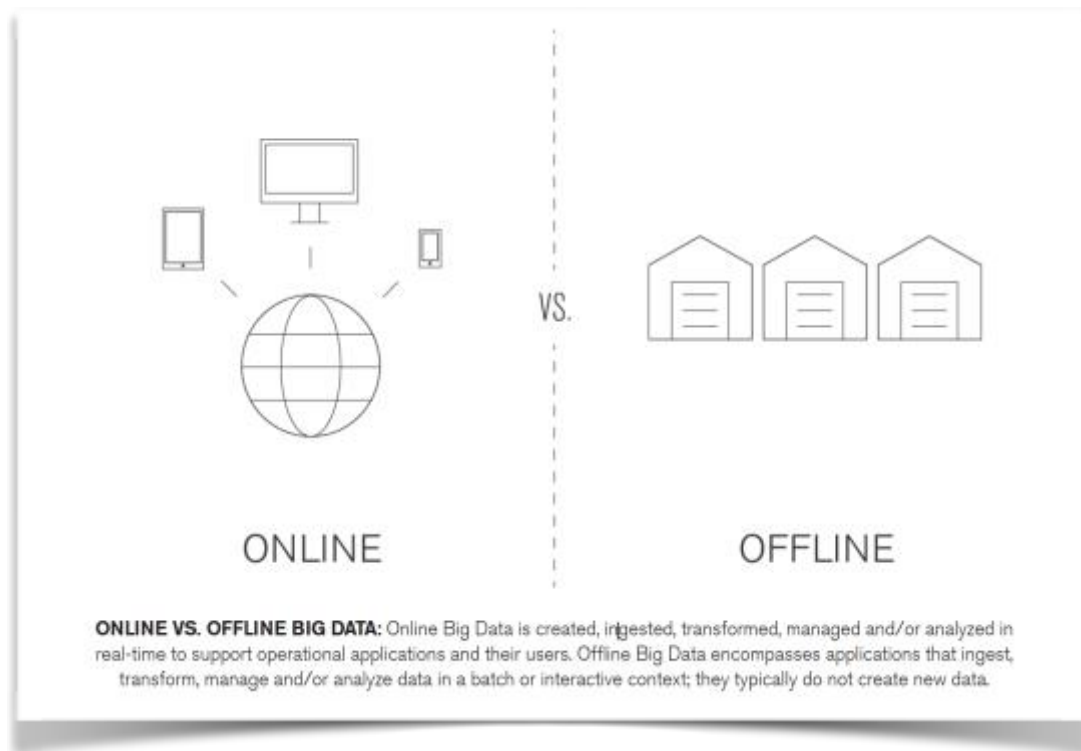


Figure 1: Online vs Offline Big Data
Source: MongoDB 2015

However, MongoDB (2015) argued that determining which Big Data technology to use is not mutually exclusive as organisations would most likely need both.

Selection Strategy for Big Data Application

Selecting the right Big Data application for an organisation requires thorough understanding of the business needs and goals. The framework in Figure 2 below, was developed to aid in understanding big data strategies and its techniques (Parise, Iyer, & Vesset 2012).

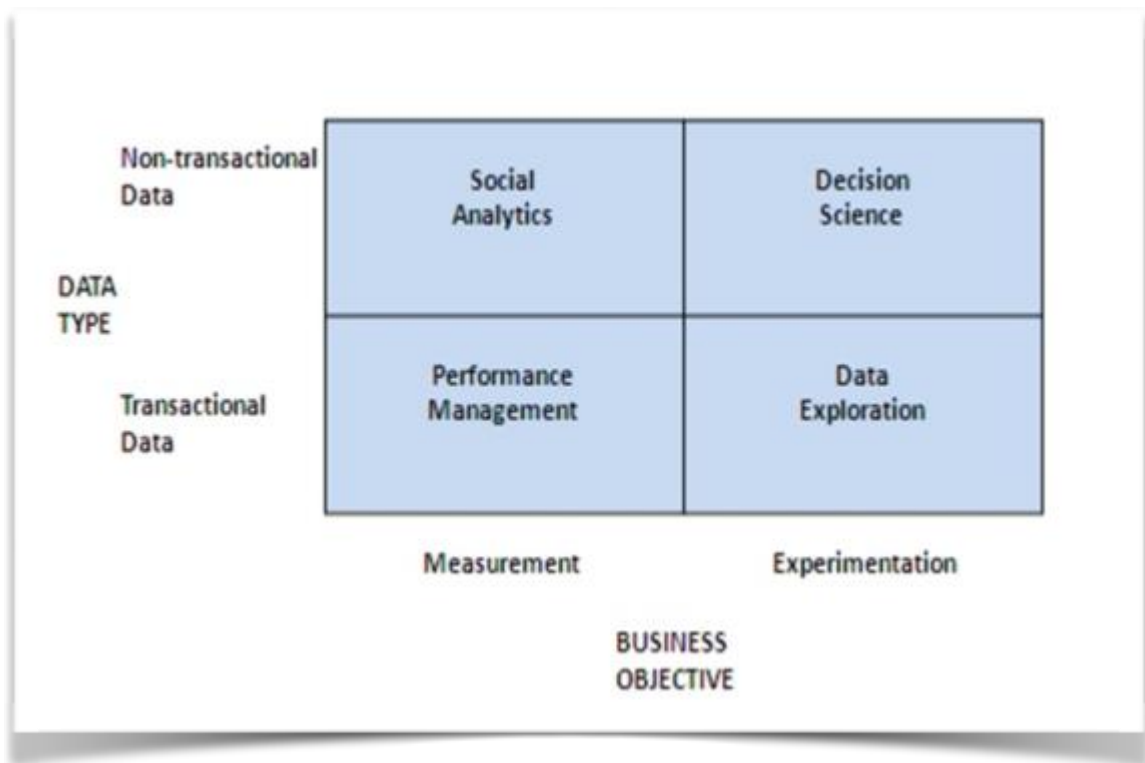


Figure 2: Big Data Framework
Source: Parise, Iyer, & Vesset 2012

Performance management, data exploration, social analytics, and decision science are the big data strategies named by Parise, Iyer, & Vesset (2012). According to the authors, in performance management, transactional data is analysed which could help in decision making. Another explanation provided by the authors is for data exploration, wherein managers make use of experimenting with statistics to get answers.

With these existing big data strategies, the authors also cited the emerging practices that go along with the strategies such as the integration of big data strategies, building a big data capability, proactivity and creating a big data policy.

Big Data Solutions Outcome

One of the results of successfully implementing big data is discovering hidden equity (Kobo 2014). The organisation explained how a correct analysis of the data could point to opportunities such as books that have high completion rates but suffer low sales.

Additionally, with Big Data, publishers could plot their marketing and publicity strategies by reading users engagement data (Kobo 2014). With this, publishers could determine if an author is worth investing on. Also, the engagement trend in a book series could easily be forecasted as a steep drop on the readers' engagement could advise a wrap up on the series.

Technologies in Big Data Solutions

To have the right knowledge, manpower and technology in Big Data is highly essential as these factors all aid in leading to accurate judgements in business decisions.

Thus, the table (Figure 3) below by Parise, Iyer, & Vesset (2012) presenting well-known big data techniques and vendors could help in deciding which technologies to use:

Transactional	Technique	Vendor			
Data	Business Intelligence (BI)/Online Analytical Processing (OLAP): <ul style="list-style-type: none">users interactively analyze multidimensional datausers can roll-up, drill-down, and slice dataBI tools provide dashboard and report capabilities	<u>Leading BI Tools:</u> Microsoft SQL Server Analysis and Reporting Services SAP BusinessObjects Oracle Business Intelligence	Non-trans- actional Social Data	Crowdsourcing: <ul style="list-style-type: none">A process for collecting data from a large community or distributed group of peopleIdea submission is a common crowdsourcing activity	<u>Leading Social Listening Tools:</u> Radian6
	Cluster Analysis: <ul style="list-style-type: none">segment objects (e.g., users) into groups based on similar properties or attributes	IBM Cognos/SPSS SAS Microstrategy		Textual Analysis: <ul style="list-style-type: none">Computer algorithms that analyze natural languageTopics can be extracted from text along with their link-ages	Attensity Visible Technologies Converseon
	Data Mining: <ul style="list-style-type: none">process to discover and extract new patterns in large data sets	QlikTech TIBCO Spotfire		Sentiment Analysis: <ul style="list-style-type: none">A form of textual analysis that determines a positive, negative, or neutral reactionOften used in marketing brand campaigns	HootSuite NodeXL network graphs
	Predictive Modeling: <ul style="list-style-type: none">a model is created to best predict the probability of an outcome				HP Autonomy
	SQL: <ul style="list-style-type: none">a computer language that manages (e.g., query, insert, delete, extract) data from a relational database			Network analysis: <ul style="list-style-type: none">A methodology to analyze the relationship among nodes (e.g., people)On social media platforms, it can be used to create the social graph of follower and friends' connections among users	Oracle Endeca IBM Watson
	A/B Testing: <ul style="list-style-type: none">A method of testing in which a control group is compared to test groups to determine if there is an improvement based on the test conditionOften used in website design to test for higher conversion rates				

Figure 3: Big Data Techniques and Vendors
Source: Ivey Business Journal

Business Impact of Big Data Solutions

The publishing field has been greatly affected by the overwhelming presence of data. With the fact that publishing has crossed its barrier from distributing containers to delivering in multiple platforms (Posth 2013, cited in Anderson 2013) shows that understanding how the Web works is vital as the future of books is headed in concept linkages (Dawson 2013, cited in Anderson 2013). But, as per McLean (2013), the industry has been lagging behind compared to other industries such as music, tv, and film in using these data for making informed decisions (McLean 2013, cited in Anderson 2013).

Thus, with all these changes and needs, publishers have embraced Big Data solutions into their organisation. Court (2015) discussed the respective actions that need to be implemented with these changes. Firstly, the author cited focus on change management as the area wherein managers should determine which departments would benefit most in analytics and solutions that come with it. Secondly, redesigning jobs was cited (Court 2015) to be part of the development since roles will be impacted when tools become available in automating tasks. Lastly, according to the author, implementing creative steps in building a culture of analytics in the organisation is necessary in order to instill a realistic action towards a data-driven culture.

On the other hand, the negative repercussions of Big Data were presented by Ganore (2012). In a blog post, the author discussed privacy issues, misuse, and unauthorized use of the data gathered. The author stated that false and overconfidence in data forecasting could cause drastic damage which could also be misused by powerful people to their advantage. Data manipulation was also cited by the author as one of the disadvantages in Big Data.

Organisational Impact of Big Data Solutions

Big data solutions are changing how organisations find profit in their services and how they provide their services. For example, sudden requests for resources and high demands in data processing were Kobo's problems during the holidays (Kobo n.d.). To solve this issue, the organization switched hosting providers to Carpathia Inc. who were able to deliver cloud computing services and cloud storage for the company.

Another example is the chart below (Figure 4) that depicts the hour that readers access the site. With this data, the newspaper could decide if they should publish at 12mn where the highest volume of readers are online (Stone 2014). The resource also discussed how the Financial Times use big data technology in customizing content recommendations for their readers.

Financial Times access by PC and mobile

Time of week and day access of FT content on mobile and print channels

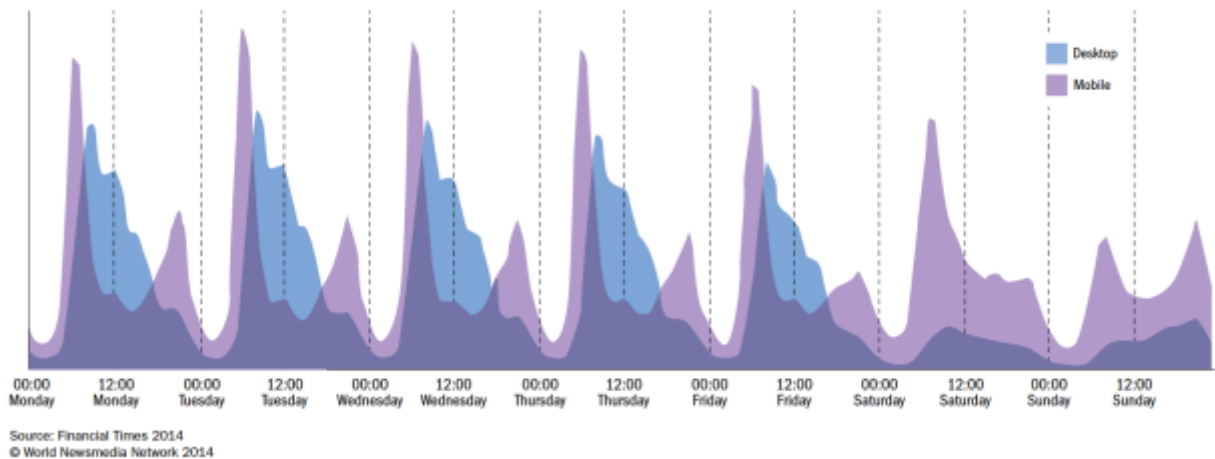


Figure 4: Financial Times access by PC and mobile
Source: Financial Times 2014

Furthermore, CNN uses big data for signs of breaking news by using a technology by Dataminr that uses an algorithm to track tweets and provide actionable alerts to clients in real-time.

Conclusion

This paper discussed the reasons Big Data technologies are essential for the publishing industry. The industry requires both real-time and long running analysis of their readers' behavior thus the technology for both online and offline data analysis is needed. With the literatures provided in this paper, it could be inferred that online data is the equivalence of non-transactional data and offline data to transactional data.

Additionally, the management of structured and unstructured data all depends on how the organisation intends to use the data. As publishing companies can be segregated into print and non-print, the use and management of unstructured data could be beneficial for non-print publishers as they could sell it to advertisers or use it for marketing purposes.

Also, with the Big Data framework presented in this study, the publishing industry could determine the factors that could aid them in decision-making and profit generation. Publishers could start with the alignment of their data type and their business objective that could lead to a specific Big Data strategy and technique custom-made for their organization.

On the other hand, Big data also presents some disadvantages. The use of Big Data technologies in analysing the readers' preferences and behaviour could also mean privacy risks and issues for the organisation and the society. Big data technologies are expensive and some companies invest large amounts of money in it, especially if the company's revenue relies largely on trend forecasting. With credibility and reliability on the line for publishing giants, organisational and societal impacts is recommended to be studied further before implementing any Big Data technology.

Overall, Big Data technology managed to change how the publishing industry used to operate. From the basic content services, delivery and management to more complex concerns of distribution, recommendation and marketing. Not only did Big Data enable publishing companies to see the analysis of readers' behaviour as a profitable area but also the trend in analysing the books' performance and authors' selling rate all became factors in the success of a publishing company. It could therefore be concluded that Big data definitely has a major place in the publishing industry and still shows it's room for growth.

References

Anderson, P 2013, *Publishing is Now a Data Game*, viewed 14 April 2016, <http://publishingperspectives.com/2013/09/publishing-is-now-a-data-game/#.Vw-Hr2ORiIN>.

Bright, G 2015, *How to Tell if Your News Media Company Really has Big Data*, viewed 11 April 2016, <http://www.inma.org/blogs/big-data-for-news-publishers/post.cfm/how-to-tell-if-your-news-media-company-really-has-big-data>.

Court, D 2015, *Getting Big Impact from Big Data*, viewed 18 April 2016, <http://www.mckinsey.com/business-functions/business-technology/our-insights/getting-big-impact-from-big-data>.

Davenport, T 2014, *Book Publishing's Big Data Future*, viewed 11 April 2016, <https://hbr.org/2014/03/book-publishings-big-data-future/>.

DigitalPublishing 101 2015, *Case Study: Working with Kobo to Merchandise eBooks*, viewed 11 April 2016, <http://digitalpublishing101.com/digital-marketing-101/ebookstore-marketing/case-study-merchandising-ebooks-with-kobo/>.

Ganore, P 2012, *Positive and Negative Impacts of Big Data*, viewed 18 April 2016, <http://www.esds.co.in/blog/positive-and-negative-impacts-of-big-data/#sthash.AL8CvEtC.dpbs>.

Innovation Enterprise 2013, *Kobo: Big Data Innovation Summit*, viewed 12 April 2016, http://www.slideshare.net/Innovation_Enterprise/jordan-christensen.

Kelly, J 2014, *Big Data: Hadoop, Business Analytics and Beyond*, viewed 11 April 2016, [http://wikibon.org/wiki/v/Big Data: Hadoop, Business Analytics and Beyond#Big Data: Real-World Use Cases](http://wikibon.org/wiki/v/Big_Data:_Hadoop,_Business_Analytics_and_Beyond#Big_Data:_Real-World_Use_Cases)

Kobo 2014, *Publishing in the Era of Big Data*, viewed 11 April 2016, <http://news.kobo.com/ir/159/20149/Publishing%20in%20the%20Era%20of%20Big%20Data%20-%20Kobo%20Whitepaper%20Fall%202014.pdf>.

MongoDB 2015, *Big Data: Examples and Guidelines for the Enterprise Decision Maker*, viewed 13 April 2016, http://s3.amazonaws.com/info-mongodb-com/10gen_Big_Data_White_Paper.pdf.

MongoDB 2015, *Online vs Offline Big Data: Selecting the Right Big Data Technology for Greater Success*, viewed 13 April 2016, <https://www.mongodb.com/scale/online-vs-offline-big-data>.

Parise, S, Iyer, B, & Vesset, D, 2012, *Four Strategies to Capture and Create Value from Big Data*, viewed 14 April 2016, <http://iveybusinessjournal.com/publication/four-strategies-to-capture-and-create-value-from-big-data/>

Press, G 2014, *12 Big Data Definitions: What's Yours?*, viewed 11 April 2016, <http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/2/#58cc49291fd7>.

Rakuten Kobo Inc. 2015, *Our Company*, viewed 11 April 2016, <https://www.kobo.com/aboutus?style=onestore&store=AU&language=en-US>.

Stone, M 2014, *Big Data for Media*, viewed 14 April 2016,
[https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Me
dia_0.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media_0.pdf)