



**CPA**

CHARTERED  
PROFESSIONAL  
ACCOUNTANTS  
CANADA

# Big Data and Business Intelligence

## TECHNOLOGY SPOTLIGHT

***“The big data revolution is not about the data. It’s about the analytics that we can come up with and that we now have to be able to understand what these data say.”***

***- Gary King, Professor and director of the Institute for Quantitative Social Science, Harvard University***

Companies need to collect, aggregate and analyze data to make better business decisions. With the help of business-intelligence tools and methodologies, companies can now analyze large volumes of data quickly and cost effectively.

In addressing the challenges of analyzing large amounts of data, companies need to know what data they have and how it can be effectively stored and subsequently accessed. Data integrity also becomes increasingly important as the reliance on data for business decisions increases. Concepts such as data-classification schemes, taxonomies and the use of meta-data should be considered. As the volume of data swells, there will be a greater need for storage and a commensurate increase in storage costs.

## Description

### Big Data

The term “big data” is more than a term used to describe “a lot of data.” Big data encapsulates tools (e.g., Hadoop, Cassandra, etc.) for processing information at high volumes, high velocity, and high variety in a way that improves insight and decision making. According to IBM, big data can be broken down into four dimensions: volume, velocity, variety, and veracity.

Businesses are confronted with the paradox of being data rich but information poor. They need more effective means of capturing the informational value this data represents. They have to find more effective means of storing, archiving, managing and retrieving that data. As more and more data is collected, organizations are seeking more effective ways to extract timely information upon which to base ever more complex decisions.

## Business Intelligence (BI)

Business intelligence encompasses the processes, tools and techniques designed to harvest insights from the large volumes of structured data within the organization. Today BI benefits from increasingly user-friendly technology and the degree to which analysis has migrated from the central IT department to the control of revenue-generating businesses. Retailers can leverage BI tools to assess how products placement will drive higher sales. For example, placing the salsa in the aisle with tortillas may result in more sales than placing it with the sauces.

## Importance

New sources of information are being developed and new techniques are needed to fully benefit from these new sources of information. The growth in unstructured information (e.g., videos, blogposts, tweets, sensor readings from the Internet of Things devices, etc.) already outpaces the traditional sources of transaction data. Big data is part of a larger trend of data-driven decision making. While organizations are keenly aware of the need to have *some* type of analytics, many organizations face challenges that include the lack of in-house expertise. Consequently, CPAs in their respective roles need to consider the applicability and potential value of investing in analytics. As mentioned above, the ultimate value proposition (i.e., insights into customers, products, services, etc.) is consistent irrespective whether it is BI or big data analytics.

## Business Benefits and Considerations

Big data can provide enormous value and benefits to organizations. These benefits include:

- increased effectiveness of business initiatives and promotions
- improved understanding of customer behaviour and market conditions
- better time-to-insight
- cost savings.

Tesco, a U.K. grocery and general merchandise retailer, links its supply chain management to weather, date and geographic data to identify and predict trends. An example of a trend the company has extracted through big data analytics is that a “16 degrees sunny Saturday in late April will cause a spike in sales. Exactly the same conditions a couple of weeks later will not, as people have had their first BBQ of the season.” In terms of overall benefits to the company, “Big Data projects deliver huge returns at Tesco; improving promotions to ensure 30% fewer gaps on shelves, predicting the weather and behaviour to deliver £6million less food wastage in the summer, £50million less stock in warehouses, optimising store operations to give £30million less wastage.”<sup>1</sup>

1 <http://cloudofdata.com/2012/10/tesco-uses-data-for-more-than-just-loyalty-cards>.

While the use of big data and BI may bring many benefits to an organization, there are some important risk areas to consider:

Risk Areas	Risk Mitigation Strategies
<p><b>Risk practitioners or other gatekeepers in the organization may opt out of big data because of general fears and uncertainties around the technologies or other doubts.</b></p>	<p>When identifying the net new risks with big data, it is important to understand how the underlying technology differs from the standard control processes around transactional systems. A good place to start is the Cloud Security Alliance's <i>Top Ten Big Data Security and Privacy Challenges</i>.<sup>2</sup> For example, the publication walks the reader through how the technology disassembles a data set and then processes the individual "chunks" of data. It points out that if the process to read the chunk of data is unauthorized then there is a risk the analytic produced will be incorrect. Those familiar with IT General Controls<sup>3</sup> will recognize this as an application development control which requires the underlying code to be tested and authorized.</p>
<p><b>Applying privacy controls that were applied to traditional database technologies may not be adequate for a big-data environment.</b></p>	<p>Complying with privacy in a big-data environment requires an understanding of how the risk profile has changed in terms of what data is now being used. For example, big-data analytics can include the use of pictures. Although facial recognition technology is in its infancy, companies, such as iOmniscient, have been advertising for a few years how their technology can be used in conjunction with big-data analytics.<sup>4</sup> Within such a context, organizations need to explore what is sufficient notice and consent to be compliant with privacy regulations such as the <i>Personal Information Protection and Electronic Documents Act</i> (PIPEDA). For example, privacy experts need to assess what is sufficient notice from a video recording perspective. Consequently, it may or may not be enough "notice" to post a sign informing customers they are being recorded and that the recording could be used for the purposes of analytics. Because Google already has the capability to search for images, it is not too much of a stretch to see how in-store cameras uploaded with Google Image Search can identify who is in the store and feed that information to sales persons. The Federal Trade Commission (FTC) has put together a best practices document that explores the privacy issues of facial recognition, citing how Kraft Foods plans to use such technology in the supermarket.<sup>5</sup> Beyond compliance, organizations need to assess the privacy sensitivities of their customers. For example, Benetton faced a backlash from its customers for planning to use RFID tags to track its inventory. Some customers feared the technology would be used to invade their privacy.<sup>6</sup></p>

2 [https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big\\_Data\\_Top\\_Ten\\_v1.pdf](https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Top_Ten_v1.pdf).

3 This is the term used by the PCAOB for "SOX Testing".

4 See [www.iomniscient.com/Media/PR/GIT2012-Face\\_Recognition\\_in\\_a\\_Crowd.pdf](http://www.iomniscient.com/Media/PR/GIT2012-Face_Recognition_in_a_Crowd.pdf). The company advertises how facial recognition can be correlated with time to see how fast cues are moving.

5 [www.ftc.gov/opa/2012/10/facialrecognition.shtm](http://www.ftc.gov/opa/2012/10/facialrecognition.shtm).

6 [www.rfidjournal.com/articles/view?471](http://www.rfidjournal.com/articles/view?471).

## Risk Areas

**Data used in the big-data analytical model is not fit for purpose or contains significant errors that would lead to erroneous decisions.**

## Risk Mitigation Strategies

Without good controls over data quality, the inclusion of “dirty data” will result in poor analysis. Ultimately, this could result in analytics that are “materially misstated”. For example, a mining company that performed analytics using poor quality data set up an oil rig at a dry well instead of a productive well, resulting in millions of wasted Euros. Therefore, it is important to verify the integrity of the data available before moving on to discussions around big-data solutions. This is explored in the publication “[A Framework for Information Integrity Controls](#)”, ensuring the integrity of information requires a multi-domain approach by exploring the controls from a content domain (e.g., accuracy of meta-data), a processing domain (e.g., ensuring accuracy of underlying program logic manipulating the information), and the information system-environment domain (e.g., logical access to the information).

An important precursor to a big-data exercise is to ensure the data has been cleansed of errors and is fit for purpose. Basic cleansing exercises should ensure fields exclude data that does not belong there (e.g., invalid states or provinces, alpha characters in numeric fields, invalid postal codes, etc.). However, trying to ensure the data is fit for purpose can be a more difficult exercise. For example, an investigation by ProPublica found software used to predict criminality was racially biased.<sup>7</sup> In other words, systemic racism was programmed into the software. In such a situation, care must be taken to ensure only non-biased data be used within the big-data predictive model.

## Conclusion

With the decreasing cost of data storage and the rising popularity of connected devices, there is no shortage of data for businesses to use. But this data needs to be cleaned, analyzed and interpreted to provide the greatest value to businesses. With the help of BI tools and knowledge of risk areas around big data, businesses can begin to apply proper analytics to their data to discover valuable insights.

7 [www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing](http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing).

This publication is part of the **Technology Spotlight series**.

The entire series covers technology trends that impact CPAs and are available on our website.

### DISCLAIMER

This paper was prepared by the Chartered Professional Accountants of Canada (CPA Canada) as non-authoritative guidance. CPA Canada and the authors do not accept any responsibility or liability that might occur directly or indirectly as a consequence of the use, application or reliance on this material.

Copyright © 2019 Chartered Professional Accountants of Canada

All rights reserved. This publication is protected by copyright. Written permission is required to reproduce, store in a retrieval system or transmit in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise).

For information regarding permission, please contact [permissions@cpacanada.ca](mailto:permissions@cpacanada.ca).