

# Big Data and Business Intelligence Tools

## TECHNOLOGY SPOTLIGHT

**Organizations are finding that traditional data-management tools are insufficient to handle the large volumes of data collected from various sources (e.g., social media, call-centre records, sensor data, data from mobile devices).**

Companies need to collect, aggregate and analyze data to make better business decisions. There are business intelligence tools and methodologies coming that will make the task doable. These tools can analyze large volumes of data in short time frames and in a cost-efficient way.

In addressing the challenges of large amounts of data, companies need to know what data they have and how it can be effectively stored and subsequently accessed. Concepts such as data-classification schemes, taxonomies and the use of metadata should be considered. As the volume of data swells, there will be a greater need for storage and commensurate increases in storage costs.

### Description

#### *Big Data*

The term “big data” is used to describe the growing accumulation of data from various internal and external sources.

Businesses are confronted with the paradox of being data rich but information poor. They are seeking more effective means of capturing the value this information represents. They are seeking more effective means to store, archive, manage and retrieve that data. With the increasing amount of data being collected, organizations are seeking more effective means to extract timely information upon which to base ever more complex decisions.

This publication was originally published by The Canadian Institute of Chartered Accountants in 2012. It has been reissued by Chartered Professional Accountants of Canada.

## **Business Intelligence**

Business intelligence encompasses the processes, tools and techniques designed to harvest critical information from the large volumes of data within the organization.

This is a long-established concept, but what makes it still relevant today is the increasing ease of the technology and the degree to which analysis has left the central IT department and moved to the control of revenue-generating businesses.

The latest versions of Microsoft's SharePoint and the success of visualization technologies such as Tableau enable the business's demand for analytical insights without the perceived delays associated with formal development methodologies.

## **Importance**

### **Big Data**

New sources of information are being developed. New techniques are needed to fully benefit from this information.

The growth in unstructured information (email, word documents, PowerPoint, etc.) already outpaces the traditional sources of transaction data.

Organizations that are able to extract useful information from the growing sources of data will have more insight into their organization, customers, suppliers, products, people and profits.

While the opportunities are significant, there are still technical obstacles.

Currently, industry observations indicate many experience few pervasive and sustainable implementations.

## **Business Intelligence**

Business intelligence traditionally affords management the opportunity to gain additional insight into the performance of an organization, and the use of that information to achieve corporate goals.

The new technologies accelerate the business's abilities to take advantage of the vast amounts of data available.

## **Business Benefits**

Business is becoming more complex. More information is required and more data is being created and stored. There is a need for:

- Appropriate taxonomies to store the data and ensure it can be efficiently and effectively retrieved;
- Tools and techniques to aid in searching databases for structured and unstructured data;
- Analysis software to provide increased insight and understanding of information from various internal and external sources;
- Effective and timely reporting tools to enable management to quickly grasp issues, risks, problems and potential solutions and make better and more timely decisions.

Issues and Risks	Possible Mitigation
<b>Big Data</b>	
<p>Few organizations know what information assets they have. Fewer still have the capability to manage those assets.</p> <p>Email and other unstructured data are rarely classified, and retention and disposition procedures are not in place or not applied.</p> <p>With e-discovery and court orders routinely applying to email, management is becoming increasingly aware of the issues associated with large volumes of data, particularly unclassified and unstructured data.</p>	<p>Organizations will benefit from an improved focus on inventorying and classifying data.</p> <p>Create a formal data architecture and data management life cycle.</p> <p>Implement a data taxonomy and standard classification and naming schemes.</p> <p>Ensure the classification and naming schemes are consistently applied to allow for correct filing and retention of data and its easy retrieval.</p>
<p>Many of the issues that we see emerging relate to future difficulties in leveraging both unstructured and structured data.</p> <p>Taxonomies exist in some industries but few are available for unstructured data.</p> <p>Many large technology providers are putting forward technical solutions; however, these generally assume the ability of the business to govern its data (e.g., having common definitions for key information assets such as customer or product). However, in many cases standard definitions do not exist or are not applied rigorously to these growing volumes of data.</p> <p>Ineffective search results or too few or too many results stemming from poor definitions or taxonomy.</p>	<p>Create and maintain an inventory of data assets.</p> <p>Ensure data assets are initially cleaned and appropriately maintained.</p> <p>Recognize that the task will only get more difficult as the database grows.</p> <p>Identify data management requirements for structured and unstructured data, as well as industry focused and industry agnostic analysis tools and techniques, particularly for the unstructured data.</p> <p>Define access controls to address privacy risks that intelligent statistical analysis of structured and unstructured data could lead to queries returning only one result (or very few results), which could allow the query user to reverse the anonymity of data/personally identifiable information that had been provided to the organization under representation of privacy (i.e. disaggregation risk).</p>
<p>The increasing need for data storage and tools to provide effective data access will drive the need to assess the suitability of storage solution.</p>	<p>Include technology and business risk as part of the assessment of the use of any technology-based or technology-reliant services.</p> <p>Ensure the business has the right to audit the service provider's security initiatives and activities; the CSP can provide a third-party audit of controls that addresses financial, operational and/or regulatory risk; and the CSP has to disclose breaches.</p>

Issues and Risks	Possible Mitigation
<i>Business Intelligence</i>	
<p>While there are enormous advantages for businesses, there are some control issues that need to be considered.</p> <p>Significant risks are associated with the quality of the insights. Quality will depend on the integrity of the data and the relevance and accuracy of the analysis. As it becomes easier to put together data from multiple sources, the risk of inappropriate aggregations increases. The lack of common data definitions will easily become an impediment.</p> <p>While business units can now more easily create analyses, multiple analyses may have the unfortunate side-effect of increasing the need for reconciliations or rationalizations.</p> <p>Two groups may even use the same data but create analyses that on the surface look inconsistent. The ability to quickly rationalize differences will be important in maintaining trust in the information presented.</p> <p>Data loss and quality entropy become more likely as control of the data moves into the hands of end-users. Typically the analysis means the data is held closer to the business and beyond the reach of the central controlled environments.</p> <p>Clumsy analysis based on non-causal relationships that randomly appear relevant may cause incorrect decisions.</p>	<p>Many of the controls associated with business-managed applications and end-user computing can be used to mitigate business intelligence handling risks.</p> <p>An inventory of where these business intelligence solutions are deployed, and the data used, ensures effective implementation.</p> <p>Evaluate the need for mitigating controls based on the risk of the various implementations.</p> <p>The ability to independently validate the analyses will be a reasonable control in cases where the risk or value of the decisions based on the analyses will require cost justification.</p> <p>In large corporations formally adopt data governance, data/information management processes and tools, and ensure enterprise-wide adoption of standard data definitions to mitigate the risk of misinterpretation or errors.</p> <p>The importance of traditional data-loss tools such as encryption and monitoring of data movements will likely become of increasing importance.</p> <p>Train staff on the use of approved analytics tools as well as ad-hoc tools to prevent errors and misuse of initial analysis.</p> <p>Create a process whereby analyses supporting key decisions is quality reviewed to validate the assumptions, basis of analyses, tools user and the reliability of the results.</p>

The matrices accompanying each Technology Spotlight are designed to create interest and awareness of some of the benefits, risks, issues and risk-mitigation strategies and techniques and are not designed to provide an exhaustive list of issues, risks or solutions. Readers are cautioned to seek professional assistance when addressing these technologies.

2012 © The Canadian Institute of Chartered Accountants