

Big Data, Bigger Opportunities for Business Management



Adhikary, James
Martinez, Noemi
Morey, Colin
Perry, Jonathan

Contents

Introduction.....	1
Key Issues	1
Management of Technology.....	1
Know Customers to Align.....	2
People	2
Leadership	2
Evaluate/Assess/Audit.....	2
Models of Big Data:.....	4
Taxonomies of Big Data Architecture:	5
Data	5
Compute Infrastructure	6
Storage Infrastructure	6
Analytics.....	6
Visualization.....	6
Security and Privacy:	7
Frameworks:	7
Zachman’s Enterprise Architecture	7
Plan of Action	8
Introduction	8
Capture and organization of data	8
Creation and linkage of advanced analytic models	8
Understanding and use of output.....	9
Critical Success Factors	9
Establishing a partnership with the business.....	9
Choosing the right technologies	9
Finding the right talent and leadership.....	9
Managing the big data lifecycle	9
Adopting best security practices	9
Resources:	11

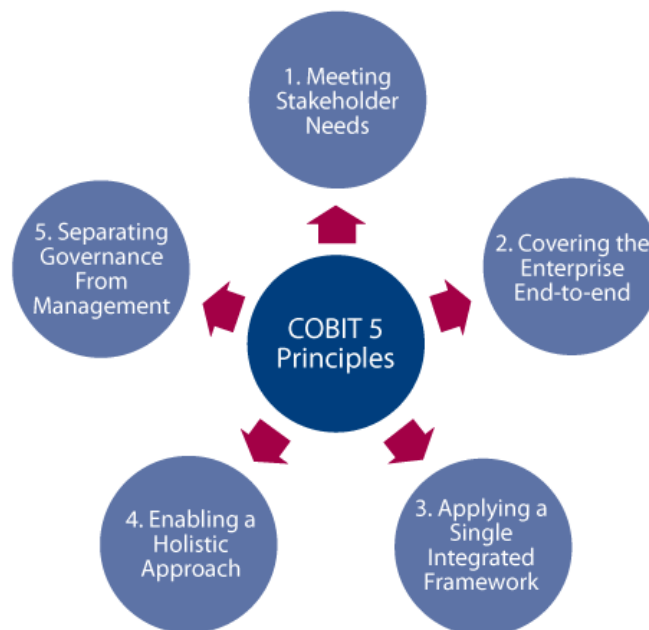
Introduction

Big data is defined by Sas.com as “a popular term used to describe the exponential growth and availability of data, both structured and unstructured”. Data is used on a regular basis, for personal and business use: from cell phones, credit cards, and computers, to the infrastructure and organization of a business. Big Data has become an important factor on how businesses collect and measure data. Online retailers can target customers more easily, by tracking what they view and suggesting products that they will buy, which helps a business have a competitive advantage over other retailers. Policemen can track where crimes will take place, which decreases violence. Governments can create better laws and health acts, which will benefit the community. With big data a lot of information can be obtained, but how can we determine which is relevant or irrelevant data? We will be looking into methods, in which business management can create value and new opportunities for their business when using big data.

Key Issues

Management of Technology

Cobit defines governance as: “Governance ensures that stakeholder needs, conditions and options are evaluated to determine balanced, agreed-on enterprise objectives to be achieved; setting direction through prioritization and decision making; and monitoring performance and compliance against agreed-on direction and objectives (Cobit.com).” Ungoverned data will generate a business a loss in revenue, decreased reliability and reputation, and diminish the value for big data analytics and operations. Governance creates the ability to discover and understand data before it is used, and provide security for both the business and customer. For governance of enterprise IT to be achieved, there must problems to be resolved and issues to address by the board and managers. Also, “the board and managers will need to accept more accountability for IT, provide guiding principles and a framework, and instill a different mindset and culture for delivering value from IT” (itgi.org).



Know Customers to Align

Big Data is all about gathering large quantities of information. Businesses have taken this opportunity to know their customers and increase profits. Big data is used by retailers to predict what a customer will buy before they even buy it, increase customer satisfaction, and be ahead of their competitors. Another industry that uses big data to know customers, is the healthcare industry. Doctors are using an individual's information (history, generics, demographic and cultural factors) to quickly and effectively diagnose their patients. Big data has changed the way businesses can conduct business, as they get to know customers better and thus provide a better service.

People

New technology is being created, that it has become difficult for IT to manage and keep up. The advances in technology, have created mass of information, or big data, that has brought innovation, productivity, and competition. The computing technology has been able to support the data, although people who have skills and talents to leverage the data have been left behind. Today, there aren't enough people with the ability to create information and outcomes out of the data. Appropriate training and education, can help business employees keep up with the technology and decrease the lack of appropriate talents. Big data is innovating the way businesses do business, but it is also creating a shortage of people who can analysis and interpret all this information.

Leadership

Leadership skills are needed in an organization to manage, organize, and make better decisions when utilizing data. Big data is not all about how much data a company carries or how much better it is, but for data to carry value it needs a strong leader who will make the best use out of that data. Leadership is about having the right representation in an organization, who can create goals, success, and opportunities. If there is no good leadership, the large quantities of structured and unstructured data will be left useless.

Evaluate/Assess/Audit

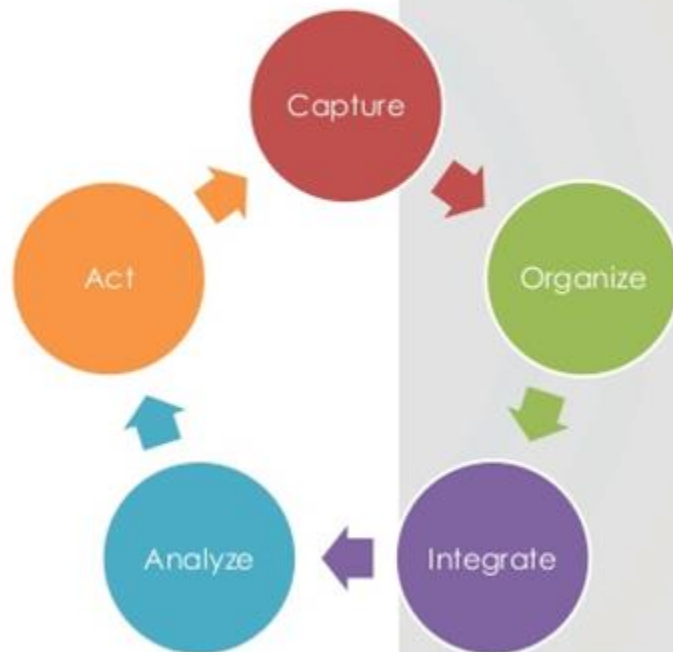
Big Data can be a success to a business if it is well managed and stored. There is so much information that a business needs to evaluate the information that is important and the information that is irrelevant to the organization. The business scorecard (shown below) can serve as guide, to help business leaders to implement the appropriate strategies needed to choose the data needed for their business. The process is to find a cost that will benefit the financial performance of the firm. One way to ensure big data doesn't affect the financial performance, is to rent out or build technologies that include data warehouse capabilities and cloud computing infrastructure in order to organize, analyze, and store large amounts of data. By doing this, the company will then have completed the "internal business process", as they are able to evaluate more efficiently the data they need. Organizational Capabilities for big data is built through constant training employees to keep up with the current technology. Then, a business will build knowledge and implement new innovation, as they keep improving the value of their data. If the organization is able to build a strong vision and strategy, they will gain customer satisfaction, and keep their shareholders happy with the revenues being produced.



Models of Big Data:

In order to build a successful big data management architecture, you must understand the cycle of big data management. You first must be able to capture your data from sensors and other sources of human generated data. Next you organize this data with a process called normalization, which is focused eliminating data redundancies and anomalies. Finally, the data can be analyzed with more accurate results based on the problem being addressed, and management can choose what action or decisions needed to make. Now that we understand the functional requirements for big data, we can take a look into the components of big data architecture.

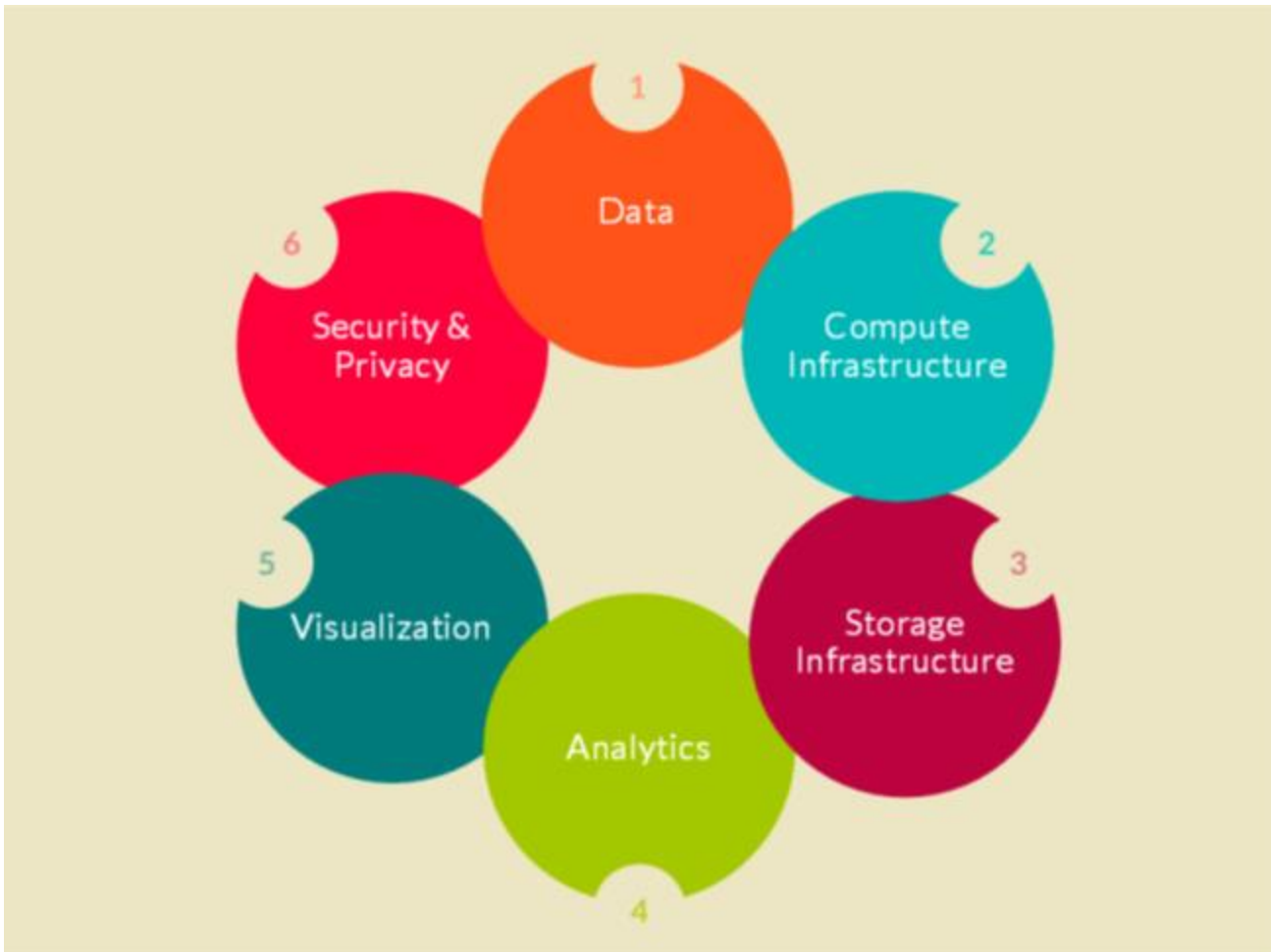
Big data cycle management



(Introduction to big data, 2014)

Taxonomies of Big Data Architecture:

To form an infrastructure for big data, you'll need to understand the 6 dimensions needed to allow for a better approach to data management, and enable your organization to optimize their data. According to Cloud Security Alliance, these 6 dimensions are:



(Big Data Taxonomy, 2014)

Data

To understand what type of infrastructure your organization is going to set up, you will first need to understand latency, and the structure of the data. Latency is the delay within a system based on delays in execution of a task. Data can be defined by the duration in which it needs to be analyze, such as in real-time, near real-time, and batch time. Real-time is defined as needing a low latency, while near real-time and batch are defined as needing average to high latency for applications that include user interaction and dashboards. Data can be structured or unstructured. Structured refers to data that has a defined length and format, while unstructured is any data that does not follow a specified format.

Compute Infrastructure

For compute infrastructure, you can use stream processing if a low latency is needed, or batch processing if low latency is not an issue. According to Sean Owen, a stream processing system is one that computes a function of one data element, or a subset of recently acquired data. Each calculation is independent and asynchronous, where the “source of data doesn’t interact with the stream processing directly” (Owen, 2015), so that processes can continue simultaneously. Sean Owen states that batch processing has access to all the data, and the main concern is with how much data can be processed in a given amount of time, making low latency not a huge issue. Once you have determined latency requirements, you can look at programming model methods such as MapReduce to approach the batch processing side, or use Apache Spark to approach the need for stream processing. Familiarizing yourself with these tools and platforms places you far ahead most companies, according to Wegene, only 38% of companies they surveyed are using the newest tools available.

Storage Infrastructure

“In order to scale databases here to handle the volume, velocity, and variety of data, we need to scale horizontally across multiple servers rather than scaling vertically by upgrading a single server (by adding more RAM or increasing HDD capacity)” (Big Data Taxonomy, 2014). The optimal solution is to use scale-out network attached storage systems, as you can add capacity in real time without crashing the whole system. There are currently three types of databases on the market currently, these are classified below as relational(SQL), NoSQL, and NewSQL and the differences between them.

	Old SQL	NoSQL	NewSQL
Relational	Yes	No	Yes
SQL	Yes	No	Yes
ACID transactions	Yes	No	Yes
Horizontal scalability	No	Yes	Yes
Performance / big volume	No	Yes	Yes
Schema-less	No	Yes	No

(Guérin, 2014)

Analytics

Big data analytics can be broken down into prescriptive, predictive, diagnostic, and descriptive to uncover any relationships or hidden patterns. Prescriptive analytics discloses what next action should be taken, while predictive analytics will tell you what events might occur. Diagnostic analytics will look at data to figure out what happened, and determine the cause. Lastly, descriptive analytics or data mining, looks at incoming data to reveal any patterns to show what is happening now. Each of these business intelligence tools can deliver substantial value to the business by aiding managers in making business decisions.

Visualization

Data visualization is a tool for viewing the context of how data is related and the impact these relationships will have on the future, giving managers the upper-hand with decision making. Visualization assists management take-in the significance of how this data gives background to the current problem being addressed

by the organization. Cloud Security Alliance defines visualization into three categories of spatial, abstract, and interactive to give powerful insights into the data.

Security and Privacy:

Data should be available only to those who have an actual need for examining or interacting with it. In order to maintain integrity of the data, organizations should perform routine audits to check I.T. controls and systems. Companies should follow ISO17799 to create of a security program that fits a company's business model, while also following legal requirements such as SOX, GLBA, and HIPPA.

Frameworks:

The conceptual framework behind creating value from big data itself includes upstream and downstream analytics consumers, and developers of processes, tools, and their approaches to the data. Upstream analytics consumers are made up of tactic decision makers and researchers who “define driving problems that require data-centric solutions employing the integrative analyses of collections of large-scale, heterogeneous data” (Payne, 2015), while your downstream analytics consumers are those “who may not necessarily define driving problems but have a research or clinical need to apply information products generated via big data analytics to advance hypothesis generation, testing, and decision making” (Payne, 2015). Developers of processes, tools, and approaches work to allow for structured and more manageable grasp of the data, assuring that stakeholders are able to understand which tools exist for big data, how the tools are adequately utilized, and how to make meaning out of the data output. The primary outcome of these components is to “make better predictions, and better predictions yield better decisions” (Brynjolfsson, E., & McAfee, A 2012).

Zachman's Enterprise Architecture

“EA is about the creation of a shared language to communicate about, think about, and manage the enterprise.” (Kappelman, 2013). We have in this report created a shared language for managers to communicate effectively with these analytical workers and developers. We took a systems thinking approach to explain the explain all aspects of the infrastructure of big data, by showing the people involved, the key issues, concepts, technology, tools, and processes to create a shared language and achieve the optimal set-up that's aligned with the business. The ontology for building a big data infrastructure like Zachman's ontology of EA was to explain the concepts and their connected relationships, so that managers can effectively implement an infrastructure for big data and to manage it.

Plan of Action

Introduction

The most important factor in effectively managing big data and putting it to good use is having a plan for management to use when collecting, compiling, and sifting through large amounts of data. However, many companies do not have an effective plan and some do not have a plan in place at all when setting out to analyze big data. The advantage of having a plan for analyzing big data is that it can highlight the essential or critical decisions a business must make and serve to enhance productivity and profitability. According to a study by McKinsey & Company states that when companies inject data and analytics deep into their operations, they can deliver productivity and profit gains that are 5 to 6 percent higher than those of the competition (Biesdorf, Court, Willmott, 2013). While many large companies use an extremely complex and involved procedure for capturing and analyzing data, a three-step approach to developing these plans or procedures can be applied to most.

These steps are:

1. The capture and organization of data
2. The creation (or purchase) and linkage of advanced analytic models
3. The understanding and use of output

Now we'll discuss these steps in more detail to show how they influence management's decisions and allow them to create tangible results for the business.

Capture and organization of data

Collecting and organization of big data and transforming it into useful business actions and decisions has become extremely important in today's business world. While this may be true, many of today's businesses have data storage models that are either vertically siloed, horizontally siloed, or in the worst case, both. In order for companies to successfully derive value from analytics, both structured and unstructured data must be collected and organized into categories that can be used by management. Management can then use these groups of data to make more informed decisions on subjects such as critical decisions, which initiatives to prioritize, which markets to enter or exit, organizational capabilities to emphasize or create, among many others.

Creation and linkage of advanced analytic models

But how do managers go about transforming this data into meaningful knowledge? The answer is the creation (or purchase) of advanced analytic models such as those offered by IBM, Microsoft, Oracle, SAP, Pentaho, Jaspersoft and others, and linking them together to provide better insights into the organization as a whole (Wayner, 2012). While these analytic models should integrate data from a wide range of organizational activities, the perfect model simply does not exist. Big data is messy, and often contains unstructured or semi-structured content (CIO, 2015). Organizations need to resist the temptation to try and capture all data, and instead focus on which metrics are needed to achieve the intended goal. With analytics, too many variables will create complexity while making the models harder to apply and maintain (Biesdorf, Court, Willmott, 2013).

Understanding and use of output

In order for the output or knowledge produced by analytics tools to assist managers in the decision making process, managers must make sure that they themselves or their frontline employees know how to use it. Data that is too complex can often overwhelm users, and can cause them to mistrust or disregard the information. Therefore, the effective use of output relies on users that competent and forward thinking when analyzing large and complex datasets, and can transform that data into better decisions for the business.

Critical Success Factors

Business success and effective management is not guaranteed with big data. The five critical success factors that must be followed in order to leverage big data are:

Establishing a partnership with the business

Establishing a partnership with the business will help simplify a lot of issues. Even though IT has been able to support the data, people within the organization lack the talent to comprehend what is going on. Big Data Analytics come from the business and it is essential for the analytics team to work in close collaboration with the business during all phases of the analytics lifecycle to ensure the ultimate success.

Choosing the right technologies

Data will be incompetent and unstructured in the big data analytics, and having the right technologies and architecture will be keen for managing multiple types of data, ranging from structured to unstructured. When selecting the right technologies, there are a lot of elements to consider. Elements such as, storage, organization, latency, availability, accessibility, integration, deployment, etc. The right technology will not only increase productivity but also help make effective management decisions.

Finding the right talent and leadership

There are generally two types of skill sets required to leverage big data, technical skills and people skills. It is essential to employ both skills. Today, there aren't enough people with the technical expertise to create information and outcomes out of the data. Businesses are trying to create a top new role, a chief analytics officer, but before they apply this position they need to solve the problem of finding the appropriate talents.

Managing the big data lifecycle

It is imperative for a business to manage its data throughout its entire lifecycle. By managing information properly over its lifetime, organizations are better equipped to deliver competitive offerings to the market faster and support business goals with less risk. The emergence of big data creates growing amounts of information, only reinforcing the need for effective data lifecycle management.

Adopting best security practices

Hardly a week goes by without a news report of another hacked database. It is best to adopt best security practices when dealing with big data because ungoverned data not only undermines consumer confidence but also generates a business a loss in revenue, decreased reliability and reputation, and diminish the value for big data analytics and operations.

Resources:

- Woods, D. (2012). Defining the Art of Big Data Leadership. Retrieved from:
<http://www.forbes.com/sites/danwoods/2012/02/10/defining-the-art-of-big-data-leadership/>
- Shaw, J. (2014). Why "Big Data" is a big deal. Retrieved from:
<http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal>
- balncscorecardreview.com. (2014). About the Balanced Scorecard. Retrieved from:
<http://www.balancedscorecardreview.com/pages/bsc-concept/>
- Big Data Taxonomy. (2014, September 1). Retrieved November 11, 2015, from
https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Taxonomy.pdf
- Brynjolfsson, E., & McAfee, A. (2012). Big Data: The Management Revolution. Retrieved November 11, 2015, from <https://hbr.org/2012/10/big-data-the-management-revolution/ar>
- Kappelman, Leon (2013). *Enterprise Architecture 101: Who, What, Where, When, Why, How?* [PowerPoint slides].
- Guérin, Laurent. *NewSQL: What's This?* Digital image. SogetiLabs. N.p., 22 Jan. 2014. Web. 08 Nov. 2015
- Introduction to big data. Digital image. (2014, August 31). Retrieved November 11, 2015, from
<http://www.slideshare.net/partagetransparents/4831-big-data-slideshare>
- Owen, S. (2015, August 3). Streaming vs Batch: The Differences | Software Engineering Daily. Retrieved November 9, 2015.
- Wegene, R., & Sinha, V. (n.d.). The value of Big Data: How analytics differentiates winners. Retrieved November 11, 2015, from http://www.bain.com/Images/BAIN_BRIEF_The_value_of_Big_Data.pdf
- Payne, P. (2015, September 21). From Theory to Practice: Generating Value from Big Data | Discovery Themes. Retrieved November 9, 2015.
- Biesdorf, Stefan, David Court, and Paul Willmott. "Big Data: What's Your Plan?" *Big Data: What's Your Plan?* 1 Mar. 2013. Web. 7 Nov. 2015. Retrieved from:
http://www.mckinsey.com/insights/business_technology/big_data_whats_your_plan
- "Strategies to Create a Big Data Management Plan: 2015 Roadmap." *CIO*. Web. 8 Nov. 2015. Retrieved from: <http://www.cio.com.au/article/563975/how-create-big-data-management-plan/>

Wayner, Peter. "7 Top Tools for Taming Big Data." *InfoWorld*. 18 Apr. 2012. Web. 8 Nov. 2015.
Retrieved from: <http://www.infoworld.com/article/2616959/big-data/7-top-tools-for-taming-big-data.html>