

Research Note

What is Big Data?

By: Devin Luco

Copyright © 2012, ASA Institute for Risk & Innovation

Keywords: Big Data, Database Management, Data Variety, Data Velocity, Data Volume, Structured Data, Unstructured Data

Abstract – This research note defines the term “Big Data” and discusses the challenges organizations are facing with the increasing amounts of data being created each day. Not only is data volume increasing at rapid rates, but also the velocity at which companies need to manage data and the variety of sources from which companies can access data are rising concerns. This note also discusses why companies should utilize Big Data and provides examples of ways companies have optimized the analysis of this data to form a competitive advantage. Finally, this note concludes with takeaways that CIOs should consider when addressing the growing trend of Big Data.

Introduction

It is no secret that technology has made tremendous strides over the past several years. As we examine the past decade, we can physically see the advancements we have made through the development and commercialization of MP3 players, smartphones, and tablets. With these new tools at our disposal, the availability and speed with which we can send and receive information is increasing at unimaginable rates. For example, who would have thought twenty years ago that we would be able to read “tweets” from our favorite football players or be able to post “status updates” about our daily lives for the world to see?

In just recent years, we have also seen advancements in the Internet through the rise of social media and search engines. However, these advancements in technology are accompanied with their economic effects and implications. We, as technology users, are sending and receiving information through mediums such as Twitter, Facebook, and Google Search with tools such as iPhones and iPads. This has flooded the databases of most organizations and has caused concern for professionals who struggle to compile, analyze, and make informed decisions from the data that is being produced. Data

scientists and engineers have coined the term “Big Data” to describe this phenomenon of large growths in data.

What is Big Data?

Although, much has been talked about Big Data in the past several years, it has now become the buzzword of 2012. So what exactly is big data? Organizations are still trying to grasp this concept, but more importantly they are trying to find ways on how to manage it.

According to Mandeep Khera, chief marketing officer of LogLogic, "most of them are concerned about big data, yet they don't understand what it means."¹ He continues, "because there's been so much said about big data, there's no clear definition and everyone is confused."² It is true that if one were to enter “Big Data” in a search engine, such as Google or Bing, there would be numerous articles struggling to explain the concept. Moreover, most of the recent articles written describe the challenges organizations face to manage Big Data but fail to thoroughly define the term. However, organizations must first understand the subject before they can successfully manage it.

Big Data refers to a collection of large and complex data sets that becomes too difficult to store, process, and manage using established database management tools.³ Traditionally, database management

tools, such as relational databases, could be used to search and query structured data to collect information in a short period of time. An example of structured data is data stored in a spreadsheet. The organization of data in a spreadsheet makes it possible to quickly search using simple algorithms or functions. However, with vast amounts of data being produced through popular technologies such as social media, the amount of unstructured data has become excessive to the point where traditional database tools are ineffective. Unstructured data does not have the same organizational data types and rules that structured data has, which results in difficulty understanding where that data is stored.⁴ Also, unstructured data is generated through audio, video, graphical, and social media messages.⁵

Characteristics of Big Data

Approximately ten years ago Meta Group analyst Doug Laney was the first to identify the increasing data management difficulties that companies were facing.⁶ Laney developed a framework, which is still used today as the industry standard, to categorize and define big data.⁷ The three characteristics that form Big Data include volume, velocity, and variety.

Risk Consultants

1. **Volume** – Currently, 2.5 billion gigabytes of data are created each day with predictions of that number doubling about every 40 months.⁸ To put this into perspective, most of the advertised personal computers available, from retailers such as Best Buy, can store between 500 and 1000 gigabytes of data. Companies, more so than ever, now have massive amounts of data at their disposal for analysis, research, and business trend predictions. For example, Walmart gathers 2.5 million gigabytes every hour from customer transactions.⁹
2. **Velocity** – This designation refers not only to the speed of data production but also to the speed of data processing in order to keep up with day-to-day operations.¹⁰ Companies that rely on real-time and near real-time data can reap the benefits of a competitive advantage if they are able to process data fast enough.¹¹ For example, through the use of real-time mobile GPS data, figuring out how many people are in the parking lot before the store is opened and the transactions are recorded can derive sales predictions.¹² This will provide analysts and managers with the real-time information they need to make informed and timely decisions.

Risk Consultants

3. **Variety** – According to Clive Longbottom, a research analyst for Quocirca, “the most important V is variety – if you cannot deal with a variety of streams coming through, you’re not doing Big Data.”¹³ Big Data takes the shape of many forms including, social media messages, updates, images, and videos.¹⁴ It also includes data collected from sensors and location signals from mobile devices.¹⁵ In this digital age, data is being generated on every topic and from every different source one can think of.

Even though this definition and categorization seems simple enough, the organizational implications still sparks confusion among in professionals.

Utilizing Big Data

The major concern of organizations regarding data is that the data has become so large and complex that it has become extremely difficult to analyze and find relevant information. Some of the questions raised by organizations include:¹⁶

1. How do we store all this data? Or, should we store all of it?
2. What is the most valuable information in this dataset?
3. How can we find valuable information without analyzing it all?
4. How can we use the information as a competitive advantage?

According to SAS, a business analytics and business software company, companies have choices on how to solve these problems. Organizations should analyze all the data they gather with the high-performance analytics tools available such as grid computing and in-database processing.¹⁷ Another solution is to determine which data is relevant on the front end by using the analytical tools that are currently available.¹⁸ This was not possible with relational database queries. In the past, relevance of data could only be determined once a query was initiated to search through stored data in warehouses.¹⁹

Although it may be overwhelming at first, there is valuable information in the large data sets that can lead to productivity and increased profits. According to a recent study conducted at the MIT Center for Digital Business, the top third industry leaders that used data to drive decisions were 5% more productive and 6% more profitable.²⁰ Some organizations are beginning to recognize this and are finding ways to utilize data to their advantage. For example, a US airlines company that was experiencing a 5- to 10- minute gap between estimated and actual arrival time in 40% of its flights used the services of a provider that collects large amounts of data on flight arrival times.²¹ The provider was able to help the airlines company bridge the

gap by developing more accurate forecasting models from the massive amounts of data it collected.

Even the federal government has recognized the benefits of utilizing Big Data. Recently, the government has launched a Big Data Research and Development Initiative to begin the development and application of Big Data technologies.²² Big Data providers, such as HP, are looking to land contracts for projects that will help the federal government get up to speed.²³ Although, the government's need to utilize Big Data is not as critical compared to the private sector, the movement towards Big Data in the private sector must be realized or else the consequences could mean losing a competitive advantage and failing altogether.

Conclusion – CIO Takeaways

Unstructured data will continue to grow rapidly as the demand and availability for sending and receiving real-time information increases. The growth will reach a point where Big Data will no longer be considered big. Instead, Big Data will become the norm²⁴ and companies will be expected to manage data that is large in volume, fast in velocity, and vast in variety. To conclude, there are a few important takeaways about Big Data that CIOs should consider:

Risk Consultants

1. **All companies should begin thinking about how to utilize Big Data** – In the early stages, only data-driven companies, such as Google and Yahoo, needed to understand user-generated data.²⁵ However, in order to stay competitive, all companies will need to be able to effectively compile, analyze, and process data to successfully cater to the needs of their customers. Analyzing data relating to customer preferences and interests can help companies better target their products.
2. **Useful data can come in all shapes and size** – remember that data worth analyzing may not always be in front of you. Server logs and sensor data can offer valuable insight by tracking customer behaviors and providing frequent data feeds.²⁶
3. **You will need a Big Data specialist** – this one is self-explanatory. If you are migrating to a Big Data platform or solution, you will need to find people that with expertise in analyzing data using these new systems.²⁷
4. **Delay worrying about organizing Big Data** – As a rule of thumb, gather the data first and then sort it out later. Unlike structured data and data warehouses, you do not need to know what you are looking for before you collect the data.²⁸

References

- ¹ Olavsrud, Thor. "Big Data Causes Concern and Big Confusion." 24 Feb. 2012. Web. 3 Dec. 2012.
<http://m.cio.com/article/700804/Big_Data_Causes_Concern_and_Big_Confusion>
- ² Olavsrud, n. pag.
- ³ Savitz, Eric. "The Death of Big Data." *Tech*. 4 Oct. 2012. Web. 3 Dec. 2012.
<<http://www.forbes.com/sites/ciocentral/2012/10/04/the-death-of-big-data/>>
- ⁴ Database Trends and Applications. "Achieving Value From Machine- and User-Generated Unstructured Data." 22 Aug. 2012. Web. 3 Dec. 2012.
<<http://www.dbta.com/Articles/Editorial/Trends-and-Applications/Achieving-Value-From--Machine--and-User-Generated--Unstructured-Data-84233.aspx>>
- ⁵ Database Trends and Applications, n. pag.
- ⁶ ITPRO. "Big Data: Is your business ready?" 29 Oct. 2012. Web. 3 Dec. 2012.
<<http://www.itpro.co.uk/643831/big-data-is-your-business-ready/2>>
- ⁷ ITPRO, n. pag.
- ⁸ McAfee, Andrew and Brynjolfsson, Erik. "Big Data: The Management Revolution." *Harvard Business Review*. 1 Oct. 2012. Web. 3 Dec. 2012.
<<http://hbr.org/2012/10/big-data-the-management-revolution/ar/1>>
- ⁹ McAfee, n. pag.
- ¹⁰ ITPRO, n. pag.
- ¹¹ McAfee, n. pag.
- ¹² McAfee, n. pag.
- ¹³ ITPRO, n. pag.
- ¹⁴ McAfee, n. pag.
- ¹⁵ McAfee, n. pag.
- ¹⁶ SAS. "Big Data – What is it?" 2012. Web. 4 Dec. 2012
<<http://www.sas.com/big-data/index.html>>
- ¹⁷ SAS, n. pag.
- ¹⁸ SAS, n. pag.
- ¹⁹ SAS, n. pag.
- ²⁰ McAfee, n. pag.
- ²¹ McAfee, n. pag.
- ²² Corbin, Kenneth. "Federal Government's Big Data Efforts Lagging." 29 Nov. 2012. Web. 4. Dec. 2012.
<http://m.cio.com/article/722603/Federal_Government_s_Big_Data_Efforts_Lagging?page=1&taxonomyId=3133>
- ²³ Corbin, Kenneth, n. pag.
- ²⁴ Savitz, Eric, n. pag.
- ²⁵ Jackson, Joab. "Five Things CIOs Should Know About Big Data." *CIO*. 12 May. 2012. Web.

4. Dec. 2012.

<http://m.cio.com/article/706290/Five_Things_CIOs_Should_Know_About_Big_Data_>

²⁶ Jackson, Joab, n. pag.

²⁷ Jackson, Joab, n. pag.

²⁸ Jackson, Joab, n. pag.