International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 3, Issue 2, August 2014)

BIG DATA
Preconditions to Productivity

Dr. Steve Hallman, DBA[1], Dr. Michel Plaisent[2], Jasur Rakhimov[3], Dr. Prosper Bernard[4]
[1,3]*MBA Program, Park University, Parkville, MO 64152 U.S.A.*
[2]*Management and Technology,* [4]*Strategy of Affairs, University of Quebec Montreal, Montreal, Canada H3C 4R2*

***Abstract*--The continuous waves of data flooding our world have created a phenomenon that has attracted businesses to leveraging this information. According to overall research the sheer volume of data generated, stored, and mined for consumer transactions insights has become economically relevant to businesses success, governmental operations, and consumer savvy transactions. The history of previous trends in IT investment and innovation and its impact on competitiveness and productivity strongly suggest that Big Data can have a similar power, namely the ability to transform our lives. The same preconditions that allowed previous waves of IT-enabled innovation to power productivity, i.e., technology innovations followed by the adoption of complementary management innovations, are in place for Big Data, vendors and suppliers of Big Data technology and advanced analytic capabilities to have at least as much ongoing impact on productivity as vendors and suppliers of other kinds of technology.**

***Keywords*—Big Data, Apache Hadoop, RFID data**

I. Introduction - Defining Big Data

Throughout the history of the rapidly changing world of technology, businesses have always taken advantage of the latest tools offered: be it a wheel, electricity, telephones, computers etc. With the creation of the Internet and the popularity of the Web 2.0 online resources, where the user is generating the data, the term Big Data has come onto the scene of business. IBM gives the following definition to the term Big Data: Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors, and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume, and variety (IBM 2013).

People should understand where the alarming velocity, volume, and variety of data come from. In order to do so, they should reflect upon the recent historical methods of data generation. Not many years ago, companies hired employees whose primary job was inputting data into an analytic computerized system. All questionnaires, applications, and surveys were manually entered into a database for analyzing.

Data is being collected at an unprecedented scale in a broad range of application areas. Decisions that previously were based on guesswork, or on painstakingly handcrafted models of reality, can now be made using data-driven mathematical models. Such Big Data analysis now drives nearly every aspect of society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences. (Jagadish, 2014, p.86)

The next stage that increased the volume of information being collected began with the introduction and wide spread usage of Web 2.0 services. These services include blogs, social media networks, and features allowing for leaving comments and reviews. The ease of data entry has significantly increased the amount of data stored. Before Web 2.0, the Internet was static. Most users would browse websites, thereby consuming information that was published by web-administrators. With Web 2.0, the user has become the key source of data entry allowing websites to become dynamic. Features such as the ability to comment, leave reviews, or even create content have become popular. Obviously, because of user-generated data entry the velocity of data income has increased.

II. The Big Data Phenomenon

A. Advantages of Big Data

The introduction of the Big Data phenomenon led to many improvements in various spheres of information technology. It is important to name and discuss some of these improvements. One of the most obvious is the technological improvement. Because the amount of data has increased tenfold, the necessity of having a platform upon which to store this data increased as well. Supply and demand has dramatically driven down the price of disk drives while at the same time increasing the capacity of storage (McCallum, 2013).

The necessity for expensive servers having the ability to process Big Data initially appeared due to the huge amount of data demanded by dynamic computer interaction. What is often unknown to consumers is that at the same time, another means of processing big data was developed – Apache Hadoop is one of them.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 3, Issue 2, August 2014)

Apache Hadoop is a software library framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models (Apache Hadoop Foundation, 2012). With the introduction of Hadoop, additional tools, under the Apache Hadoop have been widely used. Some of them include Cloudera Impala – the industry's leading massively parallel processing (MPP) SQL query engine, that runs natively in Hadoop (Impala, 2014) and Presto – an open source distributed SQL query engine for running interactive analytic queries against data sources of all sizes ranging from gigabytes to petabytes. (Presto, 2014) Apache has also introduced a real-time computation system for processing fast, large streams of data – Apache Storm. It adds reliable real-time data processing capabilities to Apache Dadoop 2.x. Storm in Hadoop helps capture new business opportunities with low-latency dashboards, security alerts, and operational enhancements integrated with other applications running in their Hadoop cluster. (Apache Storm, 2014)

New generations of Big Data Business Intelligence solutions are cost-effective and easy to use, many are available for desktop. For example, Microsoft's business intelligence tool Power BI for Excel and Office 365 can access data from the cloud, perform business modeling, collaborate with Office 365, and provide visualization capabilities in real time. This tool is available at a fraction of the cost it was five years ago. (Brands, 2014, p.65)

B. Benfits and planning

Both businesses and customers benefit from the services available only because of Big Data processing. Such services include promotions and discounts based on your interests, geographical location, and many other facts that the database determines as your profile of interests. It is very important, however, to know how to use Big Data. Biesdorf, Court, and Willmott (2013) suggest one of the ways to manage it is to make a 3-step plan.

When a plan is in place, execution becomes easier: integrating data, initiating pilot projects, and creating new tools and training efforts occur in the context of a clear vision for driving business value – a vision that's unlikely to run into funding problems or organizational opposition. Over time, the initial plan will be adjusted. Indeed, one key benefit of big data and analytics is that you can learn things about your business that you simply could not see before (Biesdorf, Court, & Willmont, 2013)."

III. Challenges Of Big Data

A. Security Protections

Aside from the advantages of Big Data, there are challenges needing to be addressed. Expensive equipment, privacy, and security issues are examples. One of the biggest downsides of Big Data is the Privacy Issue. As the collection of unstructured data becomes more economically viable, and shifts in consumer usage of technology make a much wider range of data available, there is an incentive for organizations to collect as much data as possible. Yet, just because consumers are willing to provide data does not mean that its use is free from privacy implications (Nuan & Domenico, 2013).

Another concern is security. As usage of online payments and storage of sensitive information increases, database security becomes a major concern for business and individual customers. The companies that provide access to millions of records with bank accounts, credit card numbers, or social security numbers have a very high responsibility of database protection. Fortunately, the more sophisticated the software that the company uses, the more difficult it is for hackers to access their information. Viruses, hackers, worms, trojans, phishing, malware, and many other challenges threaten the security of databases. User generated, online data often contains sensitive personal information. Such security issues have existed for as long as data storing has existed. However, it has recently become a large topic because the venues using database storage have dramatically expanded.

The disregard of security protections can manifest difficulties for consumers. If, for example, a user's account gets hacked, the hacker may use the information for personal expenditures. Many current organizations tend to focus on using technologies for building, maintaining, and accessing databases while not devoting much focus to security (Britt, 2007). The ability for an attacker to gain illegal access to a database, and compromise precious data, places not only consumers, but also organizations at great risk. Any successful techno-attacker "would have the ability to gain sensitive data such a credit card information, social security numbers, passport IDs, or other personal identification, if one were to break into a database having only light security measures" (Patterson, 2007). Because of consumer awareness stemming from security breaches such as Target's 2013 events, many organizations are under pressure to develop more sophisticated ways to avoid vulnerabilities and probable exploitation.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 3, Issue 2, August 2014)

While Big Data presents us with multiple concerns, in turn, it has been used to provide consumers, retailers, and marketers with information that has created efficiencies, enabled new capabilities, and opened up previously unrealized opportunities. Despite such clear and compelling value creation, however, Big Data is a double-edged sword that also "has the potential to erode long-term brand equity because of its tendency to cultivate a short-term decision-making mindset" (Fulgoni, 2013). In other words, because of the transparency and an easy way of finding a cheaper product, many consumers base their choice on price rather then stay loyal to one brand.

B. Core Data First

However, companies build dynamic inventories of Big Data, which most companies overlook their ability to access other existing data, instead they continuing to invest their own Big Data. The fiscal downside is that the investments fail to pay off (Ross, Beath & Quaadgras, 2013). One method to avoid such losses would be to learn to manipulate and process the data the company already has, prior to investing in more big data options. Until a company learns how to use data, analyzing it to support its operating decisions, it will not be in a position to benefit from more Big Data. That is why a good solution would be to deter the shift to Big Data, and focus on the core data first. In addition to the recommendation for a company to focus on their existing data inventory, cultural shifts melded with changes in how operations are managed are required.

The changes are:

1) Use one source of truth (in order to avoid redundancy and inaccuracy),
2) Use scorecards (in order to provide people with data about their performance),
3) Explicitly manage your business rules (follow the business plan), and
4) Use coaching and mentoring to improve performance.

(Coaching & Mentoring: Actively partner with co-workers to provide them with information, techniques, instruction, feedback and encouragement to maximize their success on the job.)

IV. Companies Utilizing Big Data

A. Usage of Big Data

Big Data becomes a valuable business tool when it is used properly. Examples of companies that benefit from Big Data are major IT companies, Social Media web sites, and online retailers. It is relatively clear why such companies benefit most (advertisement and product offers based on "your recent purchase", "items you viewed" etc.).

There are other companies, such as Proctor & Gamble and UPS, who utilize Big Data in order to improve their business.

Proctor & Gamble uses computer modeling and simulation to analyze multiple data sources – comments collected from social media, consumer sales data, RFID data, and information from the company's highly digitized processes – and makes fact-based decisions on a daily basis (Ross, et al. 2013 p.92). UPS recently installed in its vehicles and began using the data from telematics sensors merged with mapping data and other real-time reports of drop-offs and pickups from its drivers. Using this data, UPS designs routes hoping to minimize the number of left turns a driver must make to deliver a load. Such changes can generate big payoffs, because they are deployed with more than 100,000 drivers around the world (Ross, et al. 2013).

Big Data has the potential to revolutionize much more than just research. Here are some examples how it was used for other purposes:

Google's work on Google File System with MapReduce, and subsequent open source work on systems like Hadoop, have led to arguably the most extensive development and adoption of Big Data technologies, led by companies focused on the Web, such as Facebook, LinkedIn, Microsoft, Quantcast, Twitter and Yahoo! These have become the foundation for different applications, ranging from Web search to content recommendation and computational advertising. Concurrently, there have been persuasive cases made for the value of Big Data for healthcare (through home-based continuous monitoring and through integration across providers), urban planning (through fusion of high-fidelity geographical data), intelligent transportation (through analysis and visualization of live and detailed road network data), environmental modeling (through sensor networks ubiquitously collecting data), energy saving (through unveiling patterns of use), smart materials (through the new materials genome initiative), machine translation between natural languages (through analysis of large corpora), education (particularly with online courses), computational social sciences (a new methodology growing fast in popularity because of the dramatically lowered cost of obtaining data), systemic risk analysis in finance (through integrated analysis of a web of contracts to find dependencies between financial entities), homeland security (through analysis of social networks and financial transactions of possible terrorists), computer security (through analysis of logged events, known as Security Information and Event Management, or SIEM), and so on. (Jagadish, 2014, p.88)

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 3, Issue 2, August 2014)

V. Big Data: A New Competitive Advantage

A. Early adopters

The use of Big Data is becoming a crucial way for leading companies to outperform their peers. In most industries, established competitors and new entrants alike will leverage data-driven strategies to innovate, compete, and capture value. Indeed, we found early examples of such use of data in every sector we examined. In healthcare, data pioneers are analyzing the health outcomes of pharmaceuticals when they were widely prescribed, and discovering benefits and risks that were not evident during necessarily more limited clinical trials. Other early adopters of Big Data are using data from sensors embedded in products from children's toys to industrial goods to determine how these products are actually used in the real world. Such knowledge then informs the creation of new service offerings and the design of future products.

B. New Growth Opportunities

Big Data has helped to create new growth opportunities and entirely new categories of companies, such as those that combine and analyze industry data. Many of these are companies that sit in the middle of large information flows where data about products and services, buyers and suppliers, consumer preferences and intent can be captured and analyzed. Forward-thinking leaders across sectors have aggressively to built their organizations' Big Data capabilities.

In addition to the sheer scale of Big Data, the real-time and high-frequency nature of the data are also important. For example, 'now casting,' the ability to estimate metrics such as consumer confidence, immediately, something which previously could only be done retrospectively, is becoming more extensively used, adding considerable power to prediction. Similarly, the high frequency of data allows users to test theories in near real-time and to a level never before possible (McGuire, 2012).

Looking at five domains —healthcare and retail in the United States, the public sector in Europe, and manufacturing and personal location data (the location data generated by the smart mobile devices) globally some broadly applicable ways of leveraging big data emerged. (McGuire, 2012)

VI. Five Ways To Leverage Big Data

1. Big Data can unlock significant value by making information transparent. There is still a significant amount of information that is not yet captured in digital form, e.g., data that are on paper, or not made easily accessible and searchable through networks.

   It is found that up to 25 percent of the effort in some knowledge worker workgroups consists of searching for data and then transferring them to another (sometimes-virtual) location. This effort represents a significant source of inefficiency.
2. As organizations create and store more transactional data in digital form. Organizations can collect more accurate and detailed performance information on everything from product inventories to vacation / sick days and therefore expose variability and boost performance. In fact, some leading companies are using their ability to collect and analyze big data to conduct controlled experiments to make better management decisions.
3. Big Data allows ever-narrower segmentation. As well as enabling organizations to understand the needs, lifestyles and key behavioral drivers of their customers, Big data helps in making decisions with greater confidence. Furthermore, organizations implement those decisions and systematically evaluate them to ensure higher quality standards.
4. Sophisticated analytics. Sophisticated analytics can substantially improve decision-making, minimize risks, and unearth valuable insights that would otherwise remain hidden. Retailers can use algorithms to optimize decision processes such as the automatic fine-tuning of inventories and pricing in response to real-time in-store and online sales. By mining and analyzing relevant data, organizations can create and use intelligence that allows them to identify opportunities, anticipate threats, improve efficiencies, manage risk, and develop winning strategies.
5. Big Data being developed for the next generation of products and services. For instance, manufacturers are using data obtained from sensors embedded in products to create innovative after-sales service offerings such as proactive maintenance to avoid failures in new products. The number of big data applications will explode over the next few years and such as development of these applications will present paradigm challenges. This new generation of applications, including web and mobile, will be powered by big data insights in multiple industries. This will drive a huge competitive advantage enabling businesses to see new opportunities to engage with their consumers.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 3, Issue 2, August 2014)

VII. Value Created By The Use Of Big Data

If the U.S. healthcare system were to use big data creatively and effectively to drive efficiency and quality, the sector could create more than $300 billion in value every year. Two-thirds of that would be an 8 percent reduction in U.S. healthcare expenditure. In the developed economies of Europe, government administrators could create more than €100 billion in operational efficiency improvements alone by using Big Data – and that is not including employing advanced analytic tools to reduce fraud and errors and boost the collection of tax revenues (McGuire, 2012).

There are many technological issues which need to be addressed and overcome to make the best of Big Data. An example are legacy systems and incompatible operating and software standards, programs, and formats frequently prevent the integration of data and the application of the more sophisticated analytics that create worth.

VIII. Conclusion

The information presented indicates what Big Data is and what steps it went through before it became labeled as Big. It also covered both disadvantages (possible issues with security, privacy, expensive equipment, and long return of investments if the analysis is not approached in a proper way) and advantages (how companies can benefit from implementing Big Data analysis). Finally, examples included of how some companies (both IT and non-IT) utilize Big Data to gain competitive and fiscal advantages aimed at improving the business. Clearly, Big Data analysis is a very useful tool to track, improve, and plan business interventions. It is important to extract existing data from current database management systems; only after the data has been analyzed, should a company invest in further Big Data collections and analysis ventures.

Lastly, IT vendors should help organizations identify the most profitable and practical Big Data use cases and develop products and services making technologies easier to deploy, manage, and use. Embrace an open, rather than proprietary, approach, to give customers the flexibility needed to experiment with new Big Data technologies and tools.

Similarly, companies should build Big Data service from within and partner with external service providers to help businesses develop the skills they need to deploy and manage approaches. Most importantly, listen and respond to customer feedback as Big Data deployments mature and grow.

REFERENCES

[1] Apache Storm (2014) Retrieved June 2, 2014 from http://hortonworks.com/hadoop/storm/

[2] Biesdorf, S., Court, D., & Willmont, P. (2013) Big Data: What's your plan? McKinsey Quarterly, (2), 40-51

[3] Big Data to Their Advantage. Journal Of Advertising Research, 53 (4), 372-376. Doi: 10.2501/JAR-53-4-372-376

[4] Brands, K. (2014). Big Data and Business Intelligence for Managerial Accountants. *Strategic Finance, 96* (6), 64-65

[5] Cloudera Impala (2014) Retrieved June 2, 2014 from http://www.cloudera.com/content/cloudera/en/products-and-services/cdh/impala.html

[6] Fulgoni, G. (2013) Big Data: Friend or Foe of Digital Advertising? Five Ways Marketers Should Use Digital

[7] Jagadish, H. V., Gehrke, J., Labdindis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big Data and Its Technical Challenges. *Communications of the ACM, 57*(7), 86-94. Doi:10.1145/2611567

[8] McCallum, J. C., Disk Drive Prices, (2013) Retrieved January 10, 2014 from http://www.jcmit.com/diskprice.htm

[9] McGuire, T., Manyika, J, and Chui M., (2012) Why Big Data is the new competitive advantage Retrieved January 25, 2014 from http://www.iveybusinessjournal.com/topics/strategy/

[10] why-big-data-is-the-new-competitive-advantage#.U13bnqLhI6A Nuan, D & Di Domenico, M. (2013). Market research and the ethics of big data. International Journal of Market Research, 55(4), 2-13.

[11] Patterson, D. L. (2007). Data security still at risk. Communication News 44(2), 8. Retrieved March 6, 2007 from http://connection.ebscohost.com/c/articles/24036077/data-security-still-risk

[12] Presto (2014) Retrieved June 3rd, 2014 from http://prestodb.io/

[13] Ross, J. W., Beath C. M., & Quaadgras, A. (2013) You May Not Need Big Data After All. Harvard Business Review, December Vol 91(12), 90-98.

[14] What is Apache Hadoop Foundation (2012) Retrieved January 5, 2014 from http://hadoop.apache.org/

[15] What is Big Data? (2013) Retrieved January 5, 2014 from http://www.ibm.com/big-data/us/en/