The Emergence of BIG DATA in Business

2012

April 16, 2012

**Dr. M. Mahboob Rahman**

**Associate Professor**

**School of Business**

**North South University**

Subject: Submission of Term Paper for “The Emergence of BIG DATA in Business”

**Dear Sir,**

We, five group members, were assigned to do this the project for the course “BUS 516”. After a long research and thought we have decided to do our project on “The Emergence of BIG DATA in Business”.

This project is very much relevant to what we are learning in this course. Since this is a new course for us, the project has helped us a lot in understanding the practical aspects of promotional planning in company as well as in the market.

During group work we found it quite absorbing and enjoyable. As a result of which the group participation was maximum and the outcome was quite productive.

Lack of information was the main problem for us. Despite the time constraints and other problems, we have tried our best to make the term paper as logical as we could. All the group members worked hard to bring out the best and present you an absorbing and enthralling report.

Therefore, we would be obliged if you could please accept this report. If any further assistance are needed please contact us at find\_shahriar@hotmail.com.

Sincerely yours,

1. Shahriar Mustafa 103 0566 560

2. Ariful Islam Bhuiyan 111 0098 060

3. Abu Usuf Md Rashedul Amin 112 0519 060

4. Mahamud Un Nabi 121 0427 060

5. Mir Ashif Newaj    121 0844 060

**Acknowledgement letter**

At first, I would like to thank All Mighty Allah for giving us the strength and wisdom to finish the term paper. It is really a great pleasure to have this opportunity to describe the feeling of gratitude imprisoned in the core of our hearts.

From the very beginning, until the end a number of people had helped us to gather process and ultimately write the term paper. We would like to start by thanking our instructor Dr. Mahboob Rahman , Associate Professor , North South University, as without his implicit guidance and instructions; We would never be able to finish our work. He was there for us in every step of the way, and even though we made a number of mistakes, he was extremely patient with us and never reproached us for it.

We are also thankful to our family for their kind co-operation which made our task easy.

**Table of Contents**

**Letter of Transmittal ……………………………………………….……………………………….. i**

**Acknowledgement Letter ……………………………………………………..…………………. ii**

**Introduction …….……………………………………………………………………………………. 01**

**What is Big Data ..…..…………………………………………………………………..………… 01**

**Components of Big Data Processing ………….…………………………………………….. 01**

**Creating Value Through Big Data ……………………………………………………………… 03**

**A Key Basis for Companys’ competition, Growth, Productivity and Consumer Surplus ……………………………………………………….………………………………..……. 04**

**Large Number of Companies Across the Globe have Scored Early Successes in Their Use of Big Data ………………………………………………………………………………….. 05**

**Talent Shortage ………………………………………..…………………………………………….. 06**

**Big Data’s Implications in Bangladesh …………………………………………………………. 06**

**Conclusion……………………………………………………………………………………………… 07**

**Introduction**

Everyday enormous amount of data is being produced worldwide. Big Data analytics has brought a big opportunity for organizations. Companies capture trillions of bytes of information about their customers, suppliers, and operations. IT organizations are exploring the analytics technologies to explore web-based data sources and extract value from the social networking boom. In the western world, organizations are wondering about the kind of business intelligence they could derive from all the information they have at their disposal. The organizations are trying to leverage Big Data by trying to make sense from the data that they have and by securing it. In the next three to five years, there will be a widening gap between companies that understand and exploit Big Data and companies that are aware of it but do not know what to do with it. Already the forward thinking players of the banking, insurance, manufacturing, retail, wholesale, healthcare, communications, transportation, construction, utilities, and education are successfully using big data by exploiting meaningful information from all the Data they have and using those information in formulating their strategic moves. Those companies who will be able to use Big Data successfully will be clearly ahead of those who will react slowly to capitalize on Big Data.

**What is Big Data**

Big data is very large, distributed aggregations of loosely structured data that often incomplete and inaccessible. To be more specific Big data has the following characteristics stated below.

1. It works with Petabytes/exabytes of data
2. Involved million/billions of people
3. Accumulate billions/trillions of records
4. Flat schemas with a few complex interrelationships
5. Involved time-stamped events most often
6. Work out with incomplete data set
7. Includes connections between data set those are probabilistically incidental

Big Data is not only regarding cool technologies and Web 2.0 companies are testing with massive data sets. Rather it's defining new value streams based on leveraging information. (Floyer, 2011)

**Components of Big Data Processing**

Big-data projects have a number of different layers of abstraction from abstraction of the data through to running analytics against the abstracted data. Following figure shows the basic elements of analytical Big-data and their interrelationships. The higher level components help make big data projects easier and more dynamic. Hadoop is often at the center of Big-data projects, but it is not a precondition.

The components of analytical Big-data are given below

* Hadoop packaging and support organizations like **Cloudera;** to include MapReduce - essentially the compute layer of big data.
* Any File system like Hadoop Distributed File System (**HDFS**), that manages the retrieval and storing of data and metadata required for computation. Databases such as **Hbase** can also be used.
* A higher level language such as **Pig** (part of Hadoop) can be used instead of using JAVA to simplify the writing of computations.
* A data warehouse layer named **Hive** is built on top of Hadoop
* A thin Java library named **Cascading** is sits on top of Hadoop to allow suites of MapReduce jobs to be run and managed as a unit. This is a widely used as a special tool
* **CR-X**, a Semi-automated modeling tool allow to develop interactively at great speed, and can help set up the database that will run the analytics.
* **Greenplum** or **Netezza**, a specialized scale-out analytic databases allows very fast load & reload the data for the analytic models
* ISV big data analytical packages like **ClickFox** and **Merced** run against the database to help address the business issues

Hadoop is not used in Transactional Big-data projects because it is not real-time. For transactional systems that do not need a database with ACID guarantees, NoSQL databases can used, though there are constraints like weak consistency guarantees or restricting transactions to a single data item. In case of big-data transactional SQL databases that need the ACID guarantees the choices are limited. Traditional scale-up databases are naturally too costly for very large-scale deployment, and don't scale out very well. Most social medial databases have had to hand-craft solutions. Recently a new breed of scale-out SQL database have emerged with architectures that move the processing next to the data (in the same way as Hadoop), such as **Clustrix**. These allow greater scaleoutability. (Floyer, 2011)

**Creating Value through Big Data**

Day by day the data size of a particular company increasing. If the company properly analyze and utilize the data for business purpose then it can be good competitive advantages for the company that actually Google & face book currently exploit. According to Bill Hoggarth, “Big data is the new source of productivity, growth and competitive advantage”. McKinsey Global Institute conducted a research on big data where they pointed out five key areas where big data can create value like creating transparency, Employee performance improvement, segmenting populations to customer actions, Improve decision making, and innovating new product/service / business models.

1. **Creating transparency:**

If the company makes data available to the authorized person in a timely manner then it must be create transparency towards the company. In the organization it is also important to make data available to the inter-departmental use.

1. **Employee performance improvement:**

In organization it is very important to continuous improvement of the employee performance. Bid data can be important resource for the improving performance. As in data center employee’s detail work history has been recorded. So if any employee not doing the task properly then it is very easy to analyze the work history and fine the solution which will improve the performance.

1. **Segmenting populations to customize actions:**

In marketing, customer segmentation is very important. Because through this company can seize right business strategies for the customer. Big data enables firm to collect detail information and buying pattern of the customer. Through analysis if any company offers precise product & service then customer will be happier.

1. **Improve decision making:**

Big data enables firm to collect detail information about customers and competitors. So by analyzing all data set, a firm can make better decision rather than who analyze only sample information.

1. **Innovating new product/service / business models:**

Through using big data a firm can offer new product / service to the existing/new customer. Because existing customer can provide excellent suggestion for new product & service. In addition with that their customer detail history can be a good spring of new business model.

**A Key Basis for Compnays’ Competition, Growth, productivity and Consumer Surplus**

The use of Big Data is becoming a key basis for the leading organizations to outperform their competitors. “Big Data is transforming competitive opportunities in just about every industry sector including banking, insurance, manufacturing, retail, wholesale, healthcare, communications, transportation, construction, utilities, and education. It also plays important roles in retail operations such as marketing, merchandising, operations, supply chain, and new business models.” (Murnane, 2012)

“We have seen leading retailers such as the United Kingdom’s Tesco use big data to capture market share from its local competitors, and many other examples abound in industries such as financial services and insurance” (Manyika et. al., 2011). The forward thinking organizations can begin to build their big data capabilities aggressively. The efforts may take time to accomplish but clearly this will give them a competitive edge over the others. Those companies who will fail to do so will surely be left behind.

Big data is already serving a lot of organizations to increase their growth. Big Data service provider Opera Solutions has enabled a Japanese car company to assist in auctions of cars that have been repossessed or whose lease has expired. “The system breaks the car down to many component elements and sets a value to each; it also factors information about how each car sells globally and locally. The car company uses the information to set the floor price and determine the best place to auction each car. It gets this information into the hands of agents on the ground at each auction, so they have controls over the intangibles that wouldn’t show up in the data but would impact the final price, such as a car that smells like smoke.” (Denne, 2012)

Big data enables organizations to improve their efficiency and effectiveness. It allows companies to do more with less and produce higher-quality outputs, i.e., increase the value-added content of products and services. “Companies can leverage data to design products that better match customer needs. Data can even be leveraged to improve products as they are used. An example is a mobile phone that has learned its owner’s habits and preferences, that holds applications and data tailored to that particular user’s needs, and that will therefore be more valuable than a new device that is not customized to a user’s needs” (Manyika et. al., 2011).

The customers also capture a large amount of economic surplus because of big data related innovations. “The use of big data can enable improved health outcomes, higher-quality civic engagement with government, lower prices due to price transparency, and a better match between products and consumer needs” (Manyika et. al., 2011). As the use of big data spreads industry by industry, the consumers are going to get more direct and indirect benefits from it.

**Large Number of Companies Across the globe have scored early successes in their Use of Big Data**

“Wal-Mart was one of the early pioneers in this field, using predictive analytics to better identify customer preferences on a regional basis and stock their branch locations accordingly. It was an incredibly effective tactic that yielded strong ROI and allowed them to separate themselves from the retail pack” (Smith, 2011). Other industries followed the tactics of Wal-Mart to process and analyze their data and observed success.

With over 600,000 names in BluePages, IBM’s employee directory, and over 500,000 queries daily, the average search session takes two minutes. IBM needed a faster, more efficient application. Using Apache open source technologies, the IBM CIO Lab Analytics team developed a new people-search application that allows flexible queries and returns as many results as possible, as fast as possible. Additional capabilities include quick browsing and photo images. The new Faces application offers instantaneous response time, saving on average over a minute for each search session and I has also saved thousands of hours daily for IBM employees. (Applies emerging technologies to deliver instantaneous people searches, 2011)

The lab at Brigham and Women’s Hospital was looking to find a platform for computational pharmacoepidemiologic analytics that would address rapidly emerging trends. IBM Netezza partnered with the lab and installed an IBM Netezza data warehouse appliance. IBM Netezza data warehouse appliances are purpose-built to make advanced analytics on data simpler, faster and more accessible. By utilizing IBM Netezza technology in their pharmacoepidemiology research studies, the Harvard Medical School division was able to: Increase the speed of computationally-intense analysis of claims data, accelerate testing of new, more sophisticated algorithms, facilitate automation of continuous drug safety and effectiveness monitoring (Improving Pharmaceutical Research with Netezza Powered Analytics, 2012).

The Marine Institute sought to establish SmartBay as a research, test and demonstration platform for new environmental technologies—paving the way to commercialization and the development of new markets for Irish-based companies. The Institute, working with IBM, developed a pilot information system to feed environmental data into a data warehouse, where it is processed, analyzed and displayed in new ways. The project yields greater insight into the bay environment, as well as providing practical value—from understanding how water quality impacts fisheries to predicting hazard locations and more (Putting real-time data to work and providing a platform for technology development, 2010).

There are a lot of other big data success stories being added everyday. Organizations are increasingly becoming enthusiastic about using big data to gain their organizational success.

**Talent Shortage**

Lack of skills around data science is the most important barriers to adopt Big Data in the enterprise. IT experts have an outstanding chance to take benefit of training and education initiatives to polish their data skills and spot new career paths within their organizations. Data science is an art that includes statistics, operations research, math programming and data hacking

All these are not an integral part of a data scientist but the more the polished data scientist the more is the advantage in the world to benefit the organization. (Floyer, 2011)

[Watch this video of Wikibon analysts](http://www.youtube.com/watch?v=ihQV7vT2CTg) discussing the skills gap and how some organizations are [providing services to address the issue](http://servicesangle.com/blog/2011/12/05/emc-announces-new-cloud-computing-and-data-science-training-courses/). (Video links are provided at Bibliography)

**Big Data’s Implications in Bangladesh**

Since Bangladesh is focusing on its IT sector and trying to compete with other countries by digitalizing so it is the high time to think about the implication of Big Data which refers to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time.

Institutions such as public companies or defense along with other government agencies, telecom, healthcare and financial services industry need to store and process huge amount of data; so they must focus on this issue since data volume is growing up every day. For example, Bangladesh's telecom sector, which has the world's 15th largest subscriber base of around 85 million, needs to store and process a large amount of data. They have to record location data, loads of messages transaction, call records and data usage and also need to analyze the usage patterns to come up with more competitive and customized packages for different categories of users (Subramanyam, 2012). On the other hand if telecom industries need to evaluate all the subscribers’ data after next 30/35 years that would be a massive work since big data sizes are constantly moving targets currently ranging from a few dozen terabytes to many petabytes of data in a single data set (Big Data, n.d.).

Recently US President Barack Obama has announced more than USD 200 Million for Big Data research and development initiatives which includes commitments from several federal agencies to develop new technologies to manipulate and manage big quantities of data and use those technologies in science, national security, and education (Big Data’ initiative to optimise geospatial intelligence, 2012). Since Bangladesh is also a developing country and focusing more on digitalization, our government should also focus on this issue before it becomes a barriers to the development of IT sectors and other industries.

Moreover, the concept of Big Data is not at all familiar to the mass level of Bangladesh. Government should also take initiatives to create awareness along with infrastructure development. Private companies may be interested if government initiates some projects for the development of new technologies to manage big quantities of data.

**Conclusion**

As data volumes are growing exponentially, so is the concern over data preservation, access, dissemination, and usability. Many agencies has taken initiatives to research into areas such as automated analysis techniques, data mining, machine learning, privacy, and database interoperability and these will help to identify how big data can enable science in new ways and at new levels. Moreover, the science of data shall include the processes of turning data into knowledge, data mining and visualization, interoperability, search and discovery, and semantics (Big Data, n.d.). New job roles are being emerged and new technologies are being introduced to adopt converged infrastructure, shift to cloud architectures, and wrangle with massive data stores, and at the same time enterprises need to prepare the IT workforce to handle their new roles in order to achieve success.

**Bibliography**

Applies emerging technologies to deliver instantaneous people searches. (October 20, 2011) <http://www-01.ibm.com/software/success/cssdb.nsf/CS/JHUN-8MU4A8?OpenDocument&Site=corp&cty=en_us>

Big data brings big value. (February 29, 2012). *IT Web Data management.* Retrieved April 13, 2012 from <http://www.itweb.co.za/index.php?option=com_content&view=article&id=52138>

Big Data, (n.d.). Retrieved from <http://www.nitrd.gov/subcommittee/bigdata.aspx>

‘Big Data’ initiative to optimise geospatial intelligence. (10 April 2012). Retrieved from <http://www.geospatialworld.net/index.php>?

Big Data, (n.d.). Retrieved from <http://www.nitrd.gov/subcommittee/bigdata.aspx>

Denne, S. (April 6, 2012). Big Data Success Stories: Opera Solutions. The Wall Street Journal. Retrieved from <http://blogs.wsj.com/venturecapital/2012/04/06/big-data-success-stories-opera-solutions/>

Floyer,D., (2011). Enterprise Big-data. *Wikibon.* Retrieved from <http://wikibon.org/wiki/v/Enterprise_Big-data>

Improving Pharmaceutical Research with Netezza Powered Analytics, (March 15, 2012). Retrieved from <http://www-01.ibm.com/software/success/cssdb.nsf/CS/MLFX-8SD3EJ?OpenDocument&Site=corp&cty=en_us>

Manyika, J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., & Byers A. H., (2011), Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute,.* Retrieved April 13, 2012 from <http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation>

Murnane, L. G., (April 9, 2012). Big Data: The Future Is Now. Retrieved from <http://newsbreaks.infotoday.com/NewsBreaks/Big-Data-The-Future-Is-Now-81811.asp>

Putting real-time data to work and providing a platform for technology development. (December 15, 2010). Rretrieved from <http://www-01.ibm.com/software/success/cssdb.nsf/CS/GMMY-89ZE4F?OpenDocument&Site=corp&cty=en_us>

Smith, D., (2011). 5 real-world uses of big data. Retrieved from <http://gigaom.com/cloud/5-real-world-uses-of-big-data/>

Subramanyam, R., (Sunday April 1, 2012). The use of Big Data in Bangladesh context. Retrieved from <http://www.thefinancialexpress-bd.com/more.php?date=2012-04-01&news_id=125229>

**Video Links:**

<http://servicesangle.com/blog/2011/12/05/emc-announces-new-cloud-computing-and-data-science-training-courses/>

<http://www.youtube.com/watch?v=ihQV7vT2CTg>