American Journal of Advanced Research, 2018, 2–1
December. 2018, pages 35-42
doi: 10.5281/zenodo.1439344
http://www.ibii-us.org/Journals/AJAR/
ISBN 2472-9264 (Online), 2472-9256 (Print)

Big Data Adoption: Theories, Framework, Opportunities and Challenges

Ayeni Ayokunle Olusola[1, *], Faluyi Bamidele Ibitayo[2]

[1]Department of Computer Science, Federal Polytechnic, Ekiti State, Nigeria.

[2] Department of Computer Science, Federal Polytechnic, Ekiti State, Nigeria.

*Email: ayenikunle@yahoo.com

Received on May 23, 2018; revised on September 26, 2018; published on September 30, 2018

ABSTRACT

The global volume of data is exploding and many organizations rely on its accuracy to make strategic business decisions because it serves as their lifeblood and without it, they cannot function properly. Unfortunately, most decisions are based on the smaller fraction of available data i.e. structured data while the larger part, the unstructured part, is unattended. Data is rapidly expanding, changing and coming from a variety of sources, thus making the storage, protection, handling and management of both structured and unstructured data i.e. Big Data (*BD*), a challenge. This paper after comprehensively discussing the theories of data in detail, proposes a feasible, conceptual framework for Big Data adoption by any organization. It gives insight to the potentials of Big Data analytics versus Traditional Data analytics and also presents various tools and techniques that can be adopted for data analysis. Importantly, this paper would help organizations define their expectations from Big Data analytics and its influence on customer's perception. This paper also discusses the challenges and opportunities inherent in Big Data and sets a research path in solving issues arising from its analytics and adoption.

Keywords: Data, Structured Data, Unstructured Data, Big Data (BD), Big Data Analytics, Traditional Data Analytics, Framework, Opportunities, Challenges.

1 Introduction

Data is essential in any organization and the world itself is built on the foundations of data. Countries today depend on the disposition, analysis and management of data to impact lives of their citizens. The development and use of technological infrastructure is embraced to aid data generation, so that all accessible services can be enriched as they are used. As of May 2017, the world's current population was about 7.8 billion (World Population Clock, 2017) out of which 3.8 billion people are connected to the internet (World Internet Users Statistics , 2017). Access to technological infrastructures by these teaming population generates tremendous volume of data and since the 80's, the world's technological per-capita capacity to store information approximately doubles every 40 months (Hilbert M, 2011) with about 2.5 exabytes (2.5×10^{18}) of data generated every day (IBM, 2017).

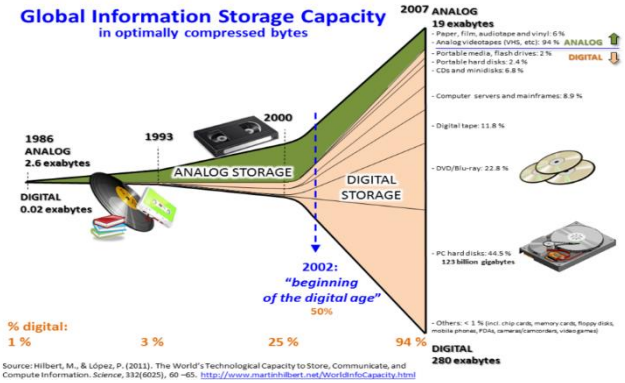

Figure 1: The World's Technological Capacity to Store, Communicate and Compute Information (Hilbert. M, 2011)

The volume of unstructured and multi-structured data within a typical organization is about 80% (Savvas, 2011) and with an av-

erage annual data growth rate of 59% (Pettey, 2011) , this percentage is bound to increase in some years. The remaining 20% is structured data and can only analyzed using traditional data analytics, leaving a high volume of valuable information source unanalyzed. The concern is not just about the volume of unstructured data posing a problem in the nearest future but also the variety and velocity are issues that should be addressed (Russom, 2011).

(Kaisler S, 2013) identified certain important issues caused by the overwhelming grsowth rate of the amount of data collected to include rapid data growth, transfer speed, diverse data, and security issues. The current trend in the improvement of data mining and storage technology however allows the preservation of this enormous volume of data.

Big data is an expression for describing large or multifaceted data sets that cannot be processed using only traditional data processing application software. Challenges posed include data capture, data storage, data analysis, data curation, search, data sharing, data transfer, data querying, data updating that needs to be processed quickly. It also refers to the use of Predictive Analytics (PA), User Behavior Analytics etc. that extract value from data, and rarely to a specific size of data set. The idea of *BD* covers unstructured, semi-structured and structured data, with the main emphasis is unstructured data (Dedić, 2017). *BD* analytics can find new ways to spot business trends, prevent diseases, combat crime and so on (Boyd, 2011).

Since Big Data is still in its early stage, this paper presents its thorough review, classifying its numerous features and then proposes a model for its adoption by any organization.

2. Background

Big data is a critical issue that requires immediate attention (Wing, 2008), (Mervis, 2012) because of the rapid growth of information put at a multiple of ten (X10) every five years (Hendrickson, 2010) which cannot be processed using existing methods and techniques (Hilbert M, 2011).

According to (James M, 2011) , many organizations will require additional investments in IT hardware, software and services to capture, store , organize and analyze large datasets since required/standard tools and techniques are not available to analyze large datasets. This causes them to experience start-up challenges in performing operations (creating, manipulating etc.) on large datasets. Replicating data also faces some security challenges regarding the reproduction of multiple data copies and policies which defines data creation, storage, analysis, relevance and access. Processing unstructured data requires understanding and addressing interests which bothers on the scalability, latency, and performance of data infrastructures (Douglas, 2012). The following subsections describes Data Measurement, Structured vs Unstructured Data, *BD* Characteristics and the growth of Demand and Supply for Storage.

2.1 Data Measurement

The binary system is used to measure digital data and its basic unit is bit ("b"), eight bits equals a byte("B"). 1kb (kilobit) or 1kB (kilobyte) of digital data equals 2^{10} and equates to 1,024 bits or 1,024 bytes, respectively. The following table demonstrates the multiplying factor related to a binary system.

Multiplying Factor	SI Prefix	Scientific Notation	Name
1 208 925 819 614 629 174 706 176	Yottabytes	2^{80}	1 septillion
1 180 591 620 717 411 303 424	Zettabytes	2^{70}	1 sextillion
1 152 921 504 606 846 976	Exabytes	2^{60}	1 quintillion
1 125 899 906 842 624	Petabytes	2^{50}	1 quadrillion
1 099 511 627 776	Terabytes	2^{40}	1 trillion
1 073 741 824	Gigabytes	2^{30}	1 billion
1 048 576	Megabytes	2^{20}	1 million
1 024	kilobytes	2^{10}	1 thousand

Table 1: Prefixes used to measure digital data using the binary system.

Since 2010, the "Zetta" prefix for digital data storage has been attained with only the "Yotta," left (Lauro, 2016), maybe it is time to add another prefix!

2.2 Structured vs unstructured data

Structured data is very organized and generally consist of tables having rows and columns that defines them e.g. relational databases.

Unstructured data is raw, disorganized and accounts for everything else. Examples include all kinds of messages (Emails, text, instant etc.), text files (.docx, .doc, PDF's etc.), presentations, audio files, video files, social media post etc. the following figures describes the terms and their growth over the past decade.

Figure 2 : Pictorial representation of Structured and Unstructured data (Shearpa, 2017).

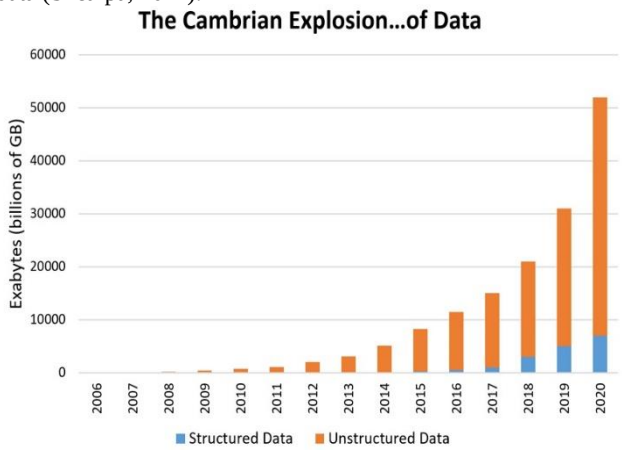

Figure 3 : Growth of structured vs unstructured data over the past decade shows that unstructured data accounts for more than 90% of all data (Lauro, 2016).

2.3 Big Data Characteristics

(Hilbert, 2015) and (Grimes, 2016) defined *BD* using the following characteristics (we call it the 5Vs):

Volume: The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered *BD* or not.

Variety: The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

Velocity: In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

Variability: Inconsistency of the data set can hamper processes to handle and manage it.

Veracity: The quality of captured data can vary greatly, affecting accurate analysis.

These characteristics are considered common issues caused by the rapid growth of *BD* with respect to data generation, analysis and management. They help discover new trends for *BD* research which has also helped us channel our strength in coming up with a useful framework for its adoption. They also help create prospects for future research in the domain of *BD*.

2.4 Growth of Demand and Supply for Storage

The demand for storage has grown more than 50% annually in recent years (James M.K, 2008), a rate faster than the ever-plunging per-gigabyte prices (Harrel, 2015) and by 2020, demand for capacity will outstrip production by six zettabytes, or nearly double the demand of 2013 alone (Lauro, 2016) The following chart depicts the rate of demand and supply for data storage:

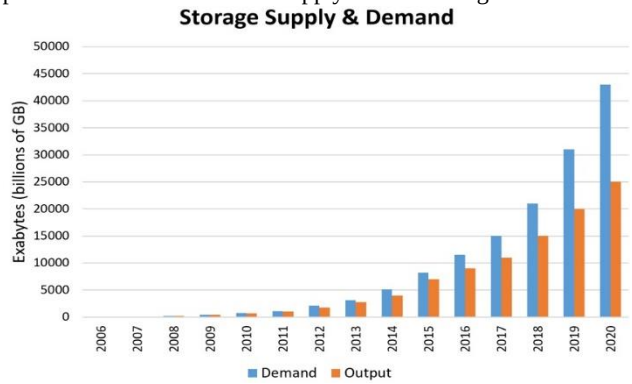

Figure 4: Storage demand and supply growth rate (Lauro, 2016)

3. Managing Big Data

Big data management is the organization, administration and governance of large volumes of both structured and unstructured data. The goal of *BD* management is to ensure a high level of data quality and accessibility for business intelligence and *BD* analytics applications (techtarget, 2013). Various sources of data hold them in disordered/messy form and until recently, most companies could not manage data using available tools within a reasonable time frame. The use of *BD* technology reduces costs and helps to validate the usefulness of *BD* to an organization before committing scare resources. When properly managed, it becomes a source of simple, dependable, secure, and manageable information and can be applied in various complex scientific disciplines complex fields including astrophysics, genomics, and biogeochemistry.

3.1 Data Storage and Management Tools

Modern computing technology provide tools and techniques for effective management of high volume of data at low cost and without sophisticated computers. These tools and techniques include Google BigTable, Simple DB, Not Only SQL (NoSQL), Data Stream Management System (DSMS) etc. (Chen M, 2014). The following describes an excellent data storage and management tool which can be used for data extraction, storage, cleaning, mining, visualizing, analyzing and integrating.

3.1.1 Hadoop

Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation (TechTarget, 2016). All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework (Apache, 2016).

Hadoop first divides files into big blocks before distributing them among nodes in a cluster. It then takes advantage of data locality by transferring packaged code into nodes for parallel processing of data (Analytics, 2017) since nodes can directly work on data they have access to. With this, datasets are processed faster and better than what is obtainable in conventional supercomputer architecture which uses parallel file system which relies on high-speed networking (Michael, 2014).

The principal part of Apache Hadoop consists of Hadoop Distributed File System (HDFS) for storage and MapReduce which serves as the processing part are inspired by Google papers on MapReduce and Google File System (EMC, 2014).

Hadoop refers not just to base modules and sub-modules (e.g. Hadoop Common, Hadoop Distributed File System (HDFS), Hadoop YARN, Hadoop MapReduce) but also to the ecosystem (Yahoo!, 2012) or group of other software packages that can be installed along with Hadoop, e.g. Apache (Pig, Hive, HBase, Phoenix, Spark, ZooKeeper, Flume, Sqoop, Oozie, Storm) and Cloudera Impala (Apache, 2016).

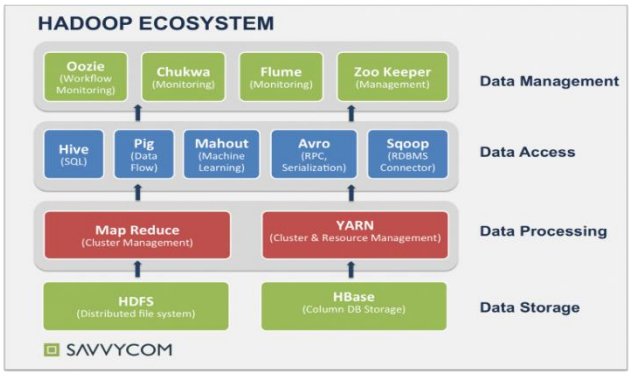

Figure 5: Hadoop ecosystem (Huydam, 2017)

3.1.2 Drawbacks of Hadoop

The major drawbacks of Hadoop includes security concerns, vulnerability to nature, not fit for small data, potential stability issues and other general limitations (Big Data Companies, 2012).

Other tools that can be considered are: cloudera, mongoDB, Talend, Refine, Quantum, Refine, Trifacta etc.

3.2 Incorporation of Big Data Solutions in an Organization's IT Setting

Hortonworker, Shaun Connolly gave a good scenario of the application (use cases) of *BD* solutions in Hortonworks industrial paper detailing three main categories of Hadoop use cases. These use cases can be generalized as types of *BD* solutions use cases, which have to be implemented in a company's current information management landscape (Connolly, 2011). They complement the traditional data systems in three possible ways:

i. Data refinery: This is the easiest infrastructural implementation of a *BD* solution. The solution is useful in analyzing large quantities of data and loading the outputs into the required data systems (to be able to access and use the extracted data with traditional tools). The traditional infrastructure system does not require any form of modification.

ii. Data exploration: data from a new *BD* solution is worked on directly, without consulting the data warehouse. This however requires good analytics application with *BD* analysis support.

iii. Application enrichment: a *BD* solution is used to directly influence an application's performance and can be referred to as the most recent/incorporated use. It is frequently used to customize user's experience in large web based companies and its implementation requires an organization's familiarity with data exploration and refinery.

4. Proposed framework for Big Data adoption and analytics testing

Successful Big Data adoption involves the combination of data with critical strategic planning and usually comes with, carry substantial consequences if it fails to meet up with expectations. This research proposes an iterative approach to develop a framework for *BD* adoption within an organization. The core components of the framework are listed below:

1. Data sourcing
2. Analytics discovery
3. Tools, Infrastructure and Technology discovery
4. Prototyping
5. Execution
6. Refinement

Unlike other models, the proposed framework adds new stages to the *BD* adoption process, i.e. prototypes and refinement stages. Since most firms in developing countries like Nigeria are cynical about *BD* solutions adoption, these stages will allow them try out *BD* platforms from *BD* firms such as Microsoft (Azure) for some months without an overwhelming contract for purchase. This allows them execute their *BD* needs and check any advantage over their current traditional analytics methods.

4.1 Framework

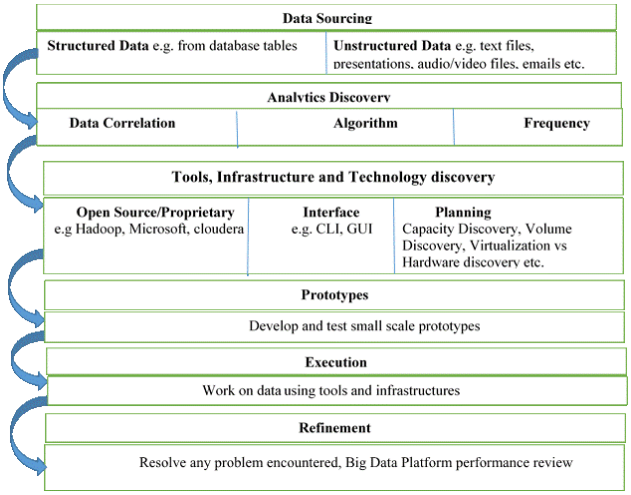

Figure 6: conceptual framework

4.1.1 Data Sourcing

Data sourcing, either by collection or generation is usually the first phase of several data lifecycles and is the first step in any *BD* adoption process. For data collection, special techniques may be required to obtain raw data from certain environments while data generation is carefully associated with the day-to-day living pattern of people. Generally, data from the Internet may not have desired values but can be accumulated over a period of time and analyzed to get useful information such as user behaviors and hobbies which can be used to predict user behaviors. The *CRM* (Customer Relationship Management System, *ERP* (Enterprise Resources Planning) platform and *OLTP* (Online Transaction Processing) systems are good sources of data.

Although, there is no globally recognized standard for storing raw data, they are transformed from their raw state and kept in desired form for easy processing.

4.1.2 Analytics Discovery

This stage deals with the following:

1. Produce a comprehensive analytics setting with special consideration for critical business functions.
2. Affirm the different sources of data and the relationship between them.
3. Produce modifiable analytics model that supports an organization's goals and can easily accommodate additional data elements.

4.1.3 Tools, Infrastructure and Technology discovery

Choosing the correct combination of tools, infrastructure and technology is crucial to a successful *BD* adoption for the following reasons:

1. The need to handle identified data sets differently. This is required to remove any interference/noise and fit for processing by the analytics engine.

2. The principal technology running *BD* analytics and other tools are still changing.

This stage is very important because it directly affects the budget of an organization and the continuous evolution of technologies and solutions makes it equally complex. It is therefore advisable to use technology components and validate their business value in parts. The server hardware, storage infrastructure and networking setup are important infrastructure decisions to be taken.

4.1.4 Prototyping
The desired *BD* technologies are tested and developed into prototypes. It also involves producing workflow for data manipulation, negotiating and exchanging data with external.

4.1.5 Execution
The adoption process is a repetitive process in which each iteration attends to a specific business area which later evolves into a comprehensive platform that caters for the entire analytical need of the organization.
The execution process adopted for the proposed framework is an iterative process which includes Project Startup, Data Identification, Solution Design, Development and Verification and fine tuning respectively.

4.1.6 Refinement
This is the final stage of the proposed framework and involves meeting up with quality assurance issues and data accuracy requirements, ensuring the selected platform meet up with business needs, solving problems encountered, making adequate recommendation etc.

4.2 Analytics Testing
The following equations were used to test the proficiency of *BD* Analytics (BDA) against Traditional Database Analytics (TDA):

BDA = Big Data Databases (MongoDB) + Unstructured Data - Man Power
Equation 4.1 BDA components

TDA = Relational Database (Mysql) + Man Power
Equation 4.2 TA components
Tests were based on the following scenarios

i. Simple database query using only SELECT statements in Mysql and its equivalent in MongoDB
ii. A complex query using INNER JOIN in SQL and
iii. A very complex query using 2 INNER JOINs with a subquery

From the scenarios stated, one can easily deduce that BDA guarantees better solution over TDA or even manpower. The main advantage of BDA is the ability to run real-time analytics when configured properly, this makes it possible to perform real time data analysis as data is received from several sources without dependence on human intervention. This reduces cost and manpower required to run BDA solutions over TDA.

5. Big Data Opportunities and Challenges
Big Data adoption presents not just several attractive opportunities but also challenges (Xiao Z, 2013). The challenges cover how data is captured, stored, shared, analyzed, visualized etc. and must be controlled to maximize the inherent opportunities in *BD*.

5.1 Challenges

5.1.1. Data mining vs Information growth
The current data mining strength cannot handle the growth of information. For example, the exploration of *BD* has been limited by computer architecture design over some years now i.e. the architectures are CPU-heavy and I/O-poor (Hey T, 2009). According to Moore's law, the performance of CPU and disk drives doubles every one and half years (18 months) (Kim NS, 2003), but the improvement has not been proportional to the rotational speed of disks over the last decade. This causes an exponential increase in sequential I/O speed whereas random I/O speed has only increased relatively.

5.1.2. Information growth vs Information processing
As earlier discussed, information grows at an exponential rate whereas, the methods and tools available for information processing are being refined to meet up with challenges at very slow pace. There are very few tools available to conveniently handle *BD* analysis and the modern techniques available in many *BD* platforms cannot effectively address the challenges of data capture, storage, analysis etc. For example, Matlab and SAS can only be used for small to medium scale data analysis; Hadoop is vulnerable to nature, not fit for small scale data processing and does not possess query processing strategies. Others have one limitation or the other, so it is safe to admit that there are not yet suitable tools to adequately explore *BD*.

5.1.3. Data inconsistency, Scalability and Security
The unpredictable nature of data in terms of inconsistency, scalability, incompleteness etc. are challenges to *BD* analysis (Labrinidis A, 2012). Since dataset are very large and obtained from heterogeneous sources in structured and unstructured form, it makes them vulnerable to inconsistency, scalability issues and security. They must be well constructed before analysis and knowing the method to use for construction is a major challenge which is required to ensure data quality and get desired result. Several preprocessing techniques must be applied to address these challenges. Consequently, it is imperative for researchers to address the lingering problems that as to do with data privacy as each stage of preprocessing is faced with its own challenges. A major challenge is deciding the appropriate encryption algorithm to encrypt large volume of data without compromising the computational power, speed and versatility.

5.1.4. Data Privacy
This is a major concern because some organizations illegally use data sourced from individuals for their personal benefit. Policies to safeguard user privacy should be put in place and audit trail must be encouraged to identify violators and apply appropriate sanctions.

5.2. Opportunities
Some of the prospects for *BD* adoption are presented below:

5.2.1. Big Data in Government
Embracing *BD* adoption in government processes ensure resourcefulness in terms of cost, productivity and novelty (Computerworld, 2012).

The USA for example keyed into this initiative when the Obama administration announced its *BD* Research and development initiative in 2012. This was targeted at knowing how *BD* could be used to address challenges faced by the government (Kalil. T,

2012). The initiative which indeed, helped the government consists of eighty-four programs distributed among six departments (Whitehouse, 2012).

BD also played a significant role in the successful re-election of President Barack Obama back to the white house in 2012 (Lampitt, 2013) and also in 2014, it helped the Bharatiya Janata Party (BJP) succeed in the Indian General Election (Livemint, 2014).

5.2.2. Global Development

A research conducted in 2012 by the United Nations Global Pulse on the use of information and communication technologies for development (ICT4D) highlighted the opportunities of utilizing new digital data sources in the field of international development (Global, 2012) and also converting large volume of data into dependable information useful in identifying and meeting the needs of low-income populations (wef, 2012).

BD analysis provides cheaper means for making important decisions in security and privacy, job creation, crime, infrastructural development, transparency and efficiency in operation and resource management (Hilbert. M, 2013), (Kvochko. E, 2012), (Medri, 2012). Data generated by users provide new opportunities to make their voice heard (Tobias. K, 2017).

5.2.3. Other Areas

This include opportunities in manufacturing, healthcare, education, media (e.g. Internet of Things), InfoTech etc.

With the widespread availability of several run-of-the-mill *BD* platforms, it is highly recommended that *BD* solutions be developed internally by companies or organizations to have a custom-tailored solution that can meet the needs of the company (Gershkoff, 2014).

6. Conclusion and further research

We have made an in depth discussion on the basic theories of Data/*BD* and successfully developed a conceptual framework by pinpointing six components for its adoption by any organization interested in implementing its solutions. Adopting *BD* solutions is a comprehensive process involving all the components identified, with each component carried out repeatedly over a period of time in order to weigh their expected business benefits against earlier assumptions as well as its feasibility in wider business strategic objectives. To help establish the required and robust *BD* solution that aids tactical decision-making process, the components identified should be carried out on a continuous basis.

BD is still at its infancy stage and thus presents several opportunities, profits, challenges and has developed such that it cannot be utilized without necessary inputs. Additional research is thus required to tackle these concerns and improve how *BD* is analyzed, displayed and stored. To support such research, human/financial resources, ground-breaking ideas etc. are required.

References

Analytics, I. (2017, May 29). *What is HDFS? - IBM Analytics*. Retrieved from IBM: https://www.ibm.com/analytics/us/en/technology/hadoop/hdfs/

Apache. (2016, October 25). *Welcome to Apache Hadoop!* Retrieved from Apache Tomcat: http://hadoop.apache.org/

Big Data Companies. (2012). *5 Big Disadvantages of Hadoop for Big Data*. Retrieved from bigdatacompanies: http://www.bigdatacompanies.com/5-big-disadvantages-of-hadoop-for-big-data/

Boyd, d. C. (2011). Six Provocations for Big Data. *Social Science Research Network: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*.

Chen M, M. S. (2014). Big data: a survey. *Mobile Networks and Applications.*, 19(2):171–209.

Cloudera. (2015, August 19). *Big Data Use Cases for Telcos.* Retrieved from Cloudera: https://www.cloudera.com/content/dam/www/marketing/resources/solution-briefs/big-data-use-cases-for-telcos.pdf.landing.html

Computerworld. (2012, 21 03). *"The Government and big data: Use, problems and potential"*. Retrieved from Computerworld: http://www.computerworld.com/article/2472667/government-it/the-government-and-big-data--use--problems-and-potential.html

Connolly, S. (2011). *Apache Hadoop Big Data Refinery White Paper*. Retrieved from hortonworks: http://info.hortonworks.com/Big-Data-Refinery-White-Paper.html

Dedić, N. S. (2017). *Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery.* Berlin: Springer International Publishing.

Douglas, K. (2012). *Infographic: big data brings marketing big numbers*. Retrieved from marketingtechblog: http://www.marketingtechblog.com/ibm-big-data-marketing/

EMC, E. S. (2014). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data.* John Wiley & Sons.

Gershkoff, A. (2014, July 11). How to Design Custom In-House BI Tools. (R. A, Interviewer)

Global, P. (2012, May). *WHITE PAPER: BIG DATA FOR DEVELOPMENT: OPPORTUNITIES & CHALLENGES (2012)*. Retrieved from United Nations Global Pulse: http://www.unglobalpulse.org/projects/BigDataforDevelopment

Grimes, S. (2016, January 5). *Big Data: Avoid 'Wanna V' Confusion*. Retrieved from informationweek: http://www.informationweek.com/big-data/big-data-

analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077?

Harrel, W. (2015, January 4). *CAN HARD DRIVE MANUFACTURERS KEEP UP WITH THE WORLD'S DEMAND?* Retrieved from digitaltrends: https://www.digitaltrends.com/computing/can-hard-drive-manufacturers-keep-worlds-demand/

Hendrickson. (2010). *Getting Started with Hadoop with Amazon's Elastic MapReduce.* EMR.

Hey T, T. S. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery.

Hilbert M, L. P. (2011). The world's technological capacity to store, communicate, and compute information. *Sciences*, 332(6025):60–65.

Hilbert, M. (2015, July 10). *Big Data for Development: A Review of Promises and Challenges. Development Policy Review.* Retrieved from martinhilbert: http://www.martinhilbert.net/big-data-for-development

Hilbert. M. (2013, January 22). *Big Data for Development: From Information- to Knowledge Societies.* Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145

Hilbert. M, L. P. (2011). *The World's Technological Capacity to Store, Communicate and Compute Information*. Retrieved from martinhilbert: http://www.martinhilbert.net/WorldInfoCapacity.html/

Huydam. (2017, January 6). *What is Hadoop Ecosystem*. Retrieved from Review Easy Homemade Cookies: http://revieweasyhomemadecookies.com/what-is-hadoop-ecosystem/

IBM. (2017). Retrieved from IBM Big Data - What is Big Data: https://www.ibm.com/big-data/us/en/

James M, M. C. (2011, May). *Big data: The next frontier for innovation, competition, and productivity.* Retrieved from mckinsey: https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiIt8HnzfLTAhXEDcAKHVo6Ab8QFgglMAA&url=https%3A%2F%2Fbigdatawg.nist.gov%2Fpdf%2FMGI_big_data_full_report.pdf&usg=AFQjCNHt7_O4tJap0rrrgE07X-r4Rp5ZBw&sig2=CkBZKiUUEe

James M.K, R. R. (2008, June). *Meeting the demand for data storage, The McKinsey Quarterly.* Retrieved from computerweekly: http://www.computerweekly.com/feature/Meeting-the-demand-for-data-storage

Jean Coumaros, S. d. (2014, September 14). *Big Data Alchemy: How can Banks Maximize the Value of their Customer Data?* Retrieved from slideshare: https://www.slideshare.net/VIRGOkonsult/bigdatainbanking-2705-v50

Kaisler S, A. F. (2013). Big data: issues and challenges moving forward. . *Proceedings of the IEEE 46th Annual Hawaii International Conference on System Sciences (HICSS '13)*, (pp. 995–1004). Hawaii.

Kaka, S. (2015). *E-GOVERNMENT ADOPTION AND FRAMEWORK FOR BIG DATA ANALYTICS IN NIGERIA.* Retrieved from covenantuniversity: http://eprints.covenantuniversity.edu.ng/5284/1/CORRECTED%20PAPER%202-E-GOVERNMENT%20ADOPTION%20IN%20NIGERIA%20AND%20FRAMEWORK%20FOR%20BIG%20DATA%20ANALYTICS.-1.pdf

Kalil. T. (2012, March 29). *https://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal*. Retrieved from whitehouse: https://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal

Kim NS, A. T. (2003). Leakage current: Moore's law meets static power. *Computer*, 36(12):68–75.

Kvochko. E. (2012, April 12). *Four Ways to Talk About Big Data.* Retrieved from The World Bank: http://blogs.worldbank.org/ic4d/four-ways-to-talk-about-big-data/

Labrinidis A, J. H. (2012). Challenges and opportunities with big data., (pp. 5(12):2032–2033).

Lampitt, A. (2013, Febuary 14). *The real story of how big data analytics helped Obama win*. Retrieved from Infoworld: http://www.infoworld.com/article/2613587/big-data/the-real-story-of-how-big-data-analytics-helped-obama-win.html

Lauro, R. (2016, September 14). *Digital Data Storage is Undergoing Mind-Boggling Growth*. Retrieved from eetimes: http://www.eetimes.com/author.asp?section_id=36&doc_id=1330462

Livemint. (2014, June 23). *Are Indian companies making enough sense of Big Data*. Retrieved from livemint: http://www.livemint.com/Industry/bUQo8xQ3gStSAy5II9lxoK/Are-Indian-companies-making-enough-sense-of-Big-Data.html

Medri, D. (2012, October 21). *Big Data & Business: An on-going revolution"*. Retrieved from Statistics View: http://www.statisticsviews.com/details/feature/5393251/Big-Data--Business-An-on-going-revolution.html

Mervis, J. (2012). Agencies rally to tackle big data. *Science*, p. 22.

Michael, M. (2014, September 19). *Data Locality: HPC vs. Hadoop vs. Spark*. Retrieved from datascienceassn: http://www.datascienceassn.org/content/data-locality-hpc-vs-hadoop-vs-spark

Olaf Acker, A. B. (2013, April 12). *Benefiting from big data: A new approach for the telecom industry.* Retrieved from pwc: https://www.strategyand.pwc.com/media/file/Strategyand_Benefiting-from-Big-Data_A-New-Approach-for-the-Telecom-Industry.pdf

Pettey, C. G. (2011). *Gartner Says Solving "Big Data" Challenge Involves More Than Just Managing Volumes of Data.* Retrieved from Gartner: http://www.gartner.com/newsroom/id/1731916

Russom, P. (2011). *ig Data Analytics. TDWI Research. SAP Further Extends Real-Time Data Platform With "Big Data" Capabilities, Orlando: MarketWatch.* Retrieved from MarketWatch.: http://www.marketwatch.com/story/sap-further-extends-real-time-data-platform-with-big-data-capabilities-2012-05-16

Savvas, A. (2011, October 25). *IBM: Businesses unable to analyse 90 percent of their*. Retrieved from ComputerworldUK.: http://www.computerworlduk.com/news/it-business/3313304/ibm-businesses-unable-to-analyse-90-percent-of-their-data/

Shearpa. (2017). *Structured and Unstructured Data: What is It?* Retrieved from shearpa software: http://sherpasoftware.com/blog/structured-and-unstructured-data-what-is-it/

techtarget. (2013, October). *What is big data management? - Definition from WhatIs.com.* Retrieved from techtarget: searchdatamanagement.techtarget.com/definition/big-data-management

TechTarget. (2016, September 15). *What is Hadoop? - Definition from WhatIs.com - SearchCloudComputing*. Retrieved from TechTarget: searchcloudcomputing.techtarget.com/definition/Hadoop

Tobias. K, J. M. (2017, January 11). *Responsible use of data*. Retrieved from dandc: https://www.dandc.eu/en/article/opportunities-and-risks-user-generated-and-automatically-compiled-data

wef. (2012, January 22). *Big Data, Big Impact: New Possibilities for International Development*. Retrieved from World Economic Forum: https://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development

Whitehouse. (2012, March). *Big data fact sheet.* Retrieved from Whitehouse: https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf

Wing, J. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 366(1881):3717–3725.

World Internet Users Statistics . (2017, May 13). Retrieved from Internet World Stats: internetworldstats.com/stats.htm

World Population Clock. (2017, May). Retrieved from worldometers: http://www.worldometers.info/world-population/

Xiao Z, X. Y. (2013). Security and privacy in cloud computing. *IEEE Communications Surveys and Tutorials*, 15(2):843–859.

Yahoo! (2012, November 14). *"Continuuity Raises $10 Million Series A Round to Ignite Big Data Application Development Within the Hadoop Ecosystem"*. Retrieved from yahoo: https://finance.yahoo.com/news/continuuity-raises-10-million-series-120500471.html