IJCSE International journal of computer sciences and engineering open access
Review Paper Volume-3, Issue-11 e-ISSN: 2347-2693

A Study on Big Data and Big Data Analytical Research and Issues

P.Kodimalar

Lecturer in Department of Computer Science,
Bharathidasan University Constituent College, Tamilnadu, India.

www.ijcseonline.org

Received: Oct/23/2015 Revised: Nov/06/2015 Accepted: Nov/18/2015 Published: Nov/27/2015

Abstract— Currently suggestions are opening to appreciate the implication of operating additional data in application to care high-quality for their strategies. It was said and proved through study cases that "More Data usually beats better algorithms". With this announcement suggestions ongoing to recognize that they be able to select to participate further in formulating superior collections of data moderately than advancing in costly algorithms. The significant volume of data is enhanced operated by way of a total the conceivable associations on a more extent, interactions that be able to certainly not be institute if the data is analyzed on separate sets or on a littler set. A greater sum of Data gives a better output but moreover working with it can become a challenge due to preparing limitations. This in this paper intends to characterize the idea of Big Data and stress the significance of Big Data Analytics.

Keywords— Big Data, Big Data Analytics, Data Acquisition, Data Generation, Data Storage

I. INTRODUCTION

Nowadays the Internet represents a huge space where great amounts of Data are included extremely day. The IBM Big Data Flood Info graphic appears that 2.7 ZB of Data exist in the digital universe today. Moreover according to this study there are 100 Terabytes updated daily through Facebook, and a lot of activity on social systems this leading to an estimate of 35 ZB of Data generated annually by 2020. Just to have an idea of the sum of Data being generated, one ZB equals 1021 bytes, meaning 1012 GB. We can associate the significance of Big Data and Big Data Investigation with the society that we live in. Today we are living in an Informational Society and we are moving towards a Information Based Society. In request to extract better information we need a greater sum of data. The Society of Data is a society where Data plays a major role in the economic, cultural and political stage. In the Information society the focused advantage is gained through understanding the Data and predicting the evolution of facts based on data. The same happens with Big Data. Extremely association needs to collect a substantial set of Data in request to support its choice and extract relationships through Data investigation as a basis for decisions. In this article we will characterize the idea of Big Data, its significance and diverse perspectives on its use. In expansion we will stress the significance of Big Data Investigation and show how the investigation of Big Data will progress decisions in the future.

II. BIG DATA CONCEPT

The term "Big Data" was first introduced to the registering world by Roger Magoulas from O'Reilly media in 2005 in request to characterize a great sum of Data that customary Data administration techniques cannot oversee and process due to the intricacy and size of this data. A study on the Evolution of Big Data as a Research and Experimental Topic appears that the term "Big Data" was present in research beginning with 1970s but has been comprised in publications in 2008. Nowadays the Big Data idea is treated from diverse focuses of view covering its implications in numerous fields. According to MiKE 2.0, the open source standard for Data Management, Big Data is characterized by its size, comprising a large, complex and independent collection of Data sets, each with the potential to interact. In addition, an imperative perspective of Big Data is the fact that it cannot be handled with standard Data administration techniques due to the inconsistency and unpredictability of the conceivable combinations. In IBM's view Big Data has four aspects: Volume: refers to the amount of Data gathered by a company.

This Data must be utilized further to obtain imperative knowledge; Velocity: refers to the time in which Big Data can be processed. Some activities are extremely imperative and need immediate responses, that is why quick preparing maximizes efficiency; Variety: Refers to the type of Data that Big Data can comprise. This Data can be organized as well as unstructured; Veracity: refers to the degree in which a leader trusts the utilized Data in request to take decision. So getting the right relationships in Big Data is extremely imperative for the business future. In addition, in Gartner's

IT Glosarry Big Data is characterized as high volume, velocity and variety Data assets that demand cost-effective, innovative forms of Data preparing for enhanced insight and choice making. According to Ed Dumbill chair at the O'Reilly Strata Conference, Big Data can be described as, "Data that exceeds the preparing capacity of conventional database systems.

The Data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain esteem from this data, you must choose an alternative way to process it." In a simpler definition we consider Big Data to be an expression that comprises diverse Data sets of extremely large, highly complex, unstructured, organized, stored and prepared utilizing particular routines and techniques utilized for business processes. There are a lot of definitions on Big Data circulating around the world, but we consider that the most imperative one is the one that each leader gives to its one company's data. The way that Big Data is characterized has implication in the framework of a business. Each leader has to characterize the idea in request to bring focused advantage for the company.

2.1 The significance of Big Data

The principle significance of Big Data comprises in the potential to progress efficiency in the content of use a substantial volume of data, of diverse type. If Big Data is characterized legitimately and utilized accordingly, associations can get a better view on their business therefore leading to efficiency in diverse areas like sales, improving the manufactured product and so forth.

Big Data can be utilized effectively in the following areas:

- In Data innovation in request to progress security and troubleshooting by analyzing the designs in the existing logs; • In client service by utilizing Data from call centers in request to get the client pattern and thus enhance client fulfillment by customizing services;

- In improving administrations and items through the use of social media content. By knowing the potential customers preferences the association can modify its product in request to address a greater area of people;

- In the detection of fraud in the online transactions for any industry; In hazard assessment by analyzing Data from the transactions on the financial market. In the future we propose to analyze the potential of Big Data and the power that can be enabled through Big Data Analysis.

2.2 Big Data challenges

The understanding of Big Data is mainly extremely important. In request to decide the best framework for a association it is essential that the Data that you are counting on must be legitimately analyzed. Moreover the time span of this investigation is imperative since some of them need to be performed extremely frequent in request to decide quickly any change in the business environment.

Another perspective is redisplayed by the new technologies that are developed extremely day. Considering the fact that Big Data is new to the associations nowadays, it is necessary for these associations to learn how to use the new developed technologies as soon as they are on the market. This is an imperative perspective that is going to bring focused advantage to a business.

The need for IT specialists it is moreover a challenge for Big Data. According to McKinsey's study on Big Data called Big Data: The next frontier for innovation, there is a need for up to 190,000 more workers with analytical expertise and 1.5 million more data- literate managers only in the United States. This Measurements are a proof that in request for an association to take the Big Data initiative has to either hire experts or train existing employees on the new field.

Security and Security are moreover imperative challenges for Big Data. Since Big Data comprises in a substantial sum of complex data, it is extremely difficult for a association to sort this Data on security levels and apply the according security. In expansion numerous of the associations nowadays are doing business cross countries and continents and the differences in security laws are considerable and have to be taken into consideration when beginning the Big Data initiative.

In our opinion for an association to get focused advantage from the manipulation of Big Data it has to take extremely good care of all factors when implementing it. One option of developing a Big Data framework is displayed below. In addition, in request to bring full abilities to Big Data each association has to take into consideration its own typical business characteristics.

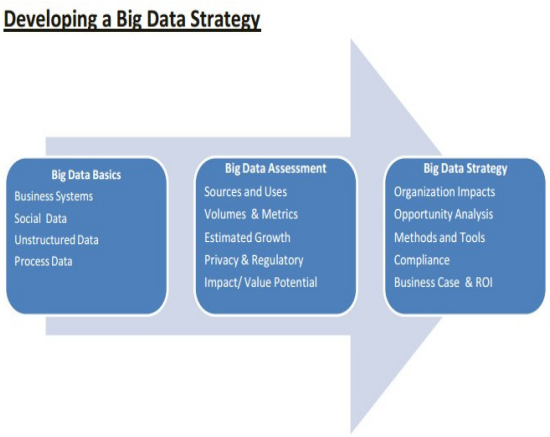

Fig 1. Developing a Big Data Strategy

III. BIG DATA ANALYTICS

The world today is constructed on the foundations of data. Lives today are impacted by the ability of the associations to dispose, interrogate and oversee data. The advancement of innovation infrastructure is adapted to help generate data, so that all the offered administrations can be improved as they are used. As an example, internet today became a big information-gathering platform due to social media and online services. At any minute they are included data. The explosion of Data cannot be any more measured in gigabytes, since Data is greater there are utilized Etabytes, Exabytes, Zettabytes and Yottabytes.

In request to oversee the giant volume of unorganized Data stored, it has been emerged the "Big Data" phenomena. It stands to reason that in the commercial sector Big-Data has been adopted more rapidly in Data driven industries, such as financial administrations and telecommunications, which it can be argued, have been experiencing a more fast growth in Data volumes compared to other market sectors, in expansion to tighter regulatory requirements and falling profitability. At first, Big Data was seen as a mean to oversee to reduce the costs of Data management. Now, the associations focus on the esteem creation potential. In request to benefit from additional insight gained there is the need to assess the analytical and execution abilities of "Big Data". To turn Big Data into a business advantage, organizations have to review the way they oversee Data inside Data center. The Data is taken from a multitude of sources, both from inside and without the organization. It can incorporate content from videos, social data, documents and machine-generated data, from a variety of applications and platforms. Organizations need a framework that is optimized for acquiring, organizing and loading this unorganized Data into their databases so that it can be effectively rendered and analyzed.

Data investigation needs to be profound and it needs to be fast and conducted with business goals in mind. The scalability of Big Data solutions inside Data centers is an essential consideration. Data is vast today, and it is only going to get bigger. If a Data center can only cope with the levels of Data expected in the short to medium term, organizations will rapidly spend on framework refreshes and upgrades. Forward planning and scalability are therefore important. In request to make extremely choice as desired there is the need to bring the results of information disco extremely to the business process and at the same time track any impact in the different dashboards, reports and exception investigation being monitored. New information discovered through investigation may moreover have a bearing on business strategy, CRM framework and financial framework going forward. See figure 2.

Fig 2. Big Data Management

Up until mid-2009 ago, the Data administration landscape was simple: Online exchange preparing (OLTP) frameworks (especially databases) bolstered the enterprise's business processes; operational Data stores (ODSs) accumulated the business transactions to support operational reporting, and enterprise Data warehouses (EDWs) accumulated and transformed business transactions to support both operational and strategic choice making. Big Data Administration is based on capturing and organizing relevant data. Data investigation assumes to understand that happened, why and predict what will happen. A deeper investigation means new analytical routines for deeper insights. Big Data investigation and the Apache Hadoop open source project are rapidly emerging as the preferred solution to business and innovation trends that are disrupting the customary Data administration and preparing landscape. Enterprises can gain a focused advantage by being early adopters of Big Data analytics.

Indeed though Big Data investigation can be technically challenging, enterprises should not delay implementation. As the Hadoop ventures mature and business intelligence (BI) tool support improves, Big Data investigation implementation intricacy will reduce, but the early adopter focused advantage will moreover wane. Innovation implementation hazard can be lessened by adapting existing

architectural principles and designs to the new innovation and changing requirements rather than rejecting them.

Big Data investigation can be differentiated from customary data-preparing architectures along a number of dimensions:-

- Speed of choice making being extremely imperative for choice makers
- Preparing intricacy since it eases the choice making process
- Transactional Data volumes which are extremely substantial
- Data structure Data can be organized and unorganized
- Flexibility of processing/investigation consisting in the sum of investigation that can be performed on it
- Concurrency

The Big Data investigation initiative should be a joint project involving both IT and business. IT should be responsible for deploying the right Big Data investigation tools and implementing sound Data administration practices. Both groups should understand that success will be measured by the esteem included by business improvements that are brought about by the initiative.

Fig 3. Prophet Big Data Solution

In terms of Big Data Administration and investigation Prophet is offering Engineered Frameworks as Big Data Solutions (Fig.3), such as Prophet Big Data Appliance, Prophet ExaData and Prophet Analytics. Big Data solutions combine best tools for each part of the problem. The customary business intelligence tools rely on relational databases for capacity and query execution and did not target Hadoop. Prophet BI combined with Prophet Big Data Connectors. The architecture assumes to load key elements of Data from Big Data sources into DBMS. Prophet Big Data connectors, Hive and Hadoop aware ETL such as ODI provide the needed Data integration capabilities. The key benefits are that the business intelligence investments and skills that are leveraged, there are made insights from Big Data consumable for business users, there are combined Big Data with Application and OLTP data.

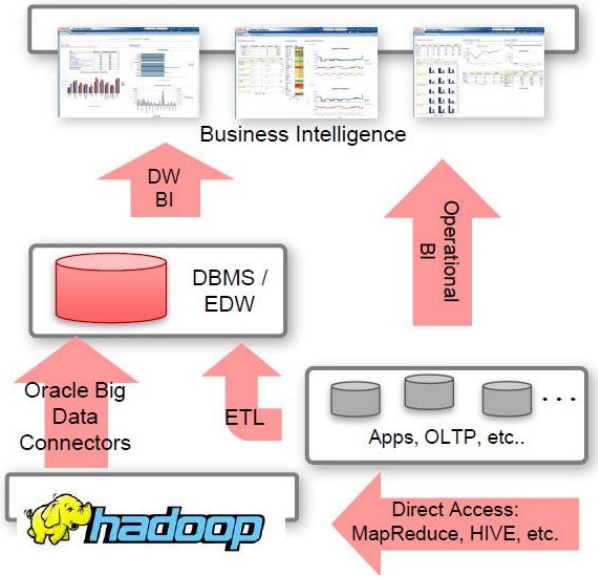

Fig 4. BI and Data Warehouse on Big Data

- Big Data provides numerous opportunities for profound insights via Data mining:
- Uncover relationships between social notion and sales data
- Predict product issues based on diagnostic sensor Data generated by items in the field
- In fact, the signal-to-noise issues often mean profound investigation to mine insight hidden in the noise is essential, as numerous forms of Big Data are simply not consumable in raw form

"Big Data" is a Data Administration & Investigation market opportunity driven by new market requirements. In-Database Investigation – Data Mining there are utilized Big Data Connectors to combine Hadoop and DBMS Data for profound analytics. Moreover there is the need to re-use SQL skills to apply deeper Data mining techniques or re-use skills for statistical analysis. Everything is all about "Big Data" instead of RAM-scale data. This is how the predictive learning of relationships between information concepts and business events is done. Big-Data presents a critical opportunity to create new esteem from giant data. It is imperative to decide appropriate governance procedures in request to oversee advancement and implementations over the life of the innovation and data. Failure to consider the longer term implications of advancement will lead to productivity issues and cost escalations. On the face of it, the cost of physically storing substantial quantities of Data

is dramatically lessened by the simplicity by which Data can be stacked into a Big-Data bunch since there is no longer required a complex ETL layer seen in any more customary Data Warehouse solutions. The bunch itself is moreover ordinarily constructed utilizing low cost product hardware and analysts are free to compose code in almost any contemporary language through the streaming API available in Hadoop.

- The business logic utilized inside an ETL flow to tokenise a stream of Data and apply Data quality standards to it must be encoded (ordinarily utilizing Java) inside each Map-Reduce program that processes the Data and any changes in source syntax or semantics

- Although the capacity hubs in a Hadoop bunch may be constructed utilizing low cost product x86 servers, the master hubs (Name Node, Auxiliary Name Hub and Job Tracker) requiring higher resilience levels to be constructed into the servers if disaster is to be avoided. Map-Reduce operations moreover generate a lot of network chatter so a quick private network is recommended. These requirements combine to add critical cost to a production bunch utilized in a commercial setting.

- Compression abilities in Hadoop are limited since of the HDFS block structure and require an understanding of the Data and compression innovation to implement adding to implementation intricacy with limited impact on capacity volumes. Other aspects to consider incorporate the true cost of ownership of pre-production and production groups such as the design build and maintenance of the groups themselves, the transition to production of Map-Reduce code to the production bunch in accordance with standard operational procedures and the advancement of these procedures. Whatever the true cost of Big-Data compared to a relational Data capacity approach, it is imperative that the advancement of Big-Data framework is consciously done, understanding the true nature of the costs and intricacy of the infrastructure, practice and procedures that are put in place.

IV. BIG DATA INVESTIGATION SOFTWARE

Currently, the trend is for enterprises to re-evaluate their approach on Data storage, administration and analytics, as the volume and intricacy of Data is growing so rapidly and unorganized Data accounting is for 90% of the Data today.

Eextremely day, 2.5 quintillion bytes of Data are made — so much that 90% of the Data in the world today has been made in the last two years alone. This Data comes from different sources such as: sensors utilized to gather climate information, posts to social media sites, digital pictures and videos, purchase exchange records, and cell phone GPS signals, web and software logs, cameras, information-sensing mobile devices, aerial sensory technologies and genomics. This Data is referred to as Big data.

"Legacy frameworks will reprinciple necessary for particular high-value, low- volume workloads, and compliment the use of Hadoop - optimizing the Data administration structure in the association by putting the right Big Data workloads in the right systems".

As it was mentioned in the Introduction Big Data compasses four dimensions: Volume, Velocity, Variety, and Veracity

Volume: Enterprises are awash with ever-growing Data of all types, easily amassing terabytes – indeed petabytes- of information(e.g. turn 12 terabytes of Tweets made each day into improved product notion analysis; convert 350 billion annual meter readings to better predict power consumption);

Velocity: For time-sensitive processes such as catching fraud, Big Data flows must be analysed and utilized as they stream into the associations in request to maximize the esteem of the information(e.g. scrutinize 5 million trade events made each day to identify potential fraud; analyze 500 million daily call detail records in constant to predict client churn faster).

Variety: Big Data comprises in any type of Data - organized and unorganized Data such as text, sensor data, audio, video, click streams, log files and more. The investigation of combined Data sorts brings new perspective for problems, situations etc.(e.g. monitor 100's of live video feeds from surveillance cameras to target focuses of interest; exploit the 80% Data growth in images, video and documents to progress client satisfaction);

Veracity: Since one of three business leaders don't trust the Data they use to make decisions, establishing trust in Big Data presents a big challenge as the variety and number of sources grows.

Apache Hadoop is a fast-growing big-Data preparing platform characterized as "an open source software project that enables the appropriated preparing of substantial Data sets over groups of product servers". It is designed to scale up from a single server to thousands of machines, with a extremely high degree of fault tolerance. Rather than depending on high-end hardware, the resiliency of these

groups comes from the software's ability to detect and handle failures at the application layer.

Developed by Doug Cutting, Cloudera's Chief Architect and the Chairman of the Apache Software Foundation, Apache Hadoop was born out of necessity as Data from the web exploded, and grew far beyond the ability of customary frameworks to handle it. Hadoop was initially inspired by papers distributed by Google outlining its approach to handling an avalanche of data, and has since become the de facto standard for storing, preparing and analyzing hundreds of terabytes, and indeed petabytes of data. Apache Hadoop is 100% open source, and pioneered a fundamentally new way of storing and preparing data. Instead of depending on expensive, proprietary hardware and diverse frameworks to store and process data, Hadoop enables appropriated parallel preparing of big amounts of Data over inexpensive, industry-standard servers that both store and process the data, and can scale without limits.

In today's hyper-connected world where more and more Data is being made extremely day, Hadoop's breakthrough advantages mean that organizations and associations can now find esteem in Data that was recently considered useless.

Hadoop can handle all sorts of Data from disparate systems: structured, unstructured, log files, pictures, audio files, communications records, email - regardless of its native format. Indeed when diverse sorts of Data have been stored in unrelated systems, it is conceivable to store it all into Hadoop bunch with no prior need for a schema. By making all Data useable, Hadoop provides the support to decide inedited relationships and reveal answers that have always been just out of reach.

In addition, Hadoop's cost advantages over legacy frameworks characterize the economics of data. Legacy systems, while fine for certain workloads, simply were not engineered with the needs of Big Data in mind and are far too costly to be utilized for general purpose with today's largest Data sets.

Apache Hadoop has two principle subprojects:

- MapReduce - The framework that understands and assigns work to the hubs in a cluster. Has been characterized by Google in 2004 and is able to distribute Data workloads over thousands of nodes. It is based on "break issue up into littler sub-problems" framework and can be exposed via SQL and in SQL-based BI tools;

- Hadoop Appropriated File Framework (HDFS) - An Apache open source appropriated file framework that compasses all the hubs in a Hadoop bunch for Data storage. It links together the file frameworks on numerous local hubs to make them into one Big file system. HDFS assumes hubs will fail, so it achieves reliability by imitating Data over multiple nodes.

HDFS is expected to run on high- performance product hardware; it is known for highly scalable capacity and automatic Data replication over three hubs for fault tolerance. Furthermore, automatic Data replication over three hubs eliminates need for backup (compose once, read numerous times). Hadoop is supplemented by an eco framework of Apache projects, such as Pig, Hive and Zookeeper that extend the esteem of Hadoop and progress its usability.

Due to the cost-effectiveness, scalability and streamlined architectures, Hadoop changes the economics and the dynamics of substantial scale computing, having a remarkable influence based on four salient characteristics.

Hadoop enables a registering solution that is:

Scalable: New hubs can be included as needed, and included without needing to change Data formats, how Data is loaded, how jobs are written, or the applications on top.

Cost effective: Hadoop brings massively parallel registering to product servers. The result is a sizeable decrease in the cost per terabyte of storage, which in turn makes it affordable to model all your data.

Flexible: Hadoop is schema-less, and can absorb any type of data, organized or not, from any number of sources. Data from multiple sources can be joined and aggregated in arbitrary ways enabling deeper analyses than any one framework can provide.

Fault tolerant: When you lose a node, the framework redirects work to another area of the Data and continues preparing without missing a beat.

Content mining makes sense of text- rich Data such as insurance claims, warranty claims, client surveys, or the growing streams of client comments on social networks. Optimization helps retailers and consumer goods makers, among others, with tasks such as setting costs for the best conceivable balance of strong-yet- profitable sales. Forecasting is utilized by insurance companies, for example, to estimate exposure or losses in the event of a hurricane or flood.

Cost will certainly be a software selection factor as that's a Big reason associations are adopting Hadoop; they're trying to retain and make use of all their data, and they're expecting cost savings over conventional relational

databases when scaling out over hundreds of Terabytes or more. Sears, for example, has more than 2 petabytes of Data on hand, and until it implemented Hadoop two years ago, Shelley says the association was constantly outgrowing databases and still couldn't store everything on one platform.

Once the application can run on Hadoop it will presumably be able to handle ventures with indeed greater and more varied Data sets, and clients will be able to rapidly analyze new Data sets without the delays associated with transforming Data to meet a rigid, pre-characterized Data model as required in relational environments. From architectural point of view, Hadoop comprises of the Hadoop Common which provides access to the file-frameworks bolstered by Hadoop.

The Hadoop Common bundle contains the necessary JAR files and scripts needed to start Hadoop. The bundle moreover provides source code, documentation, and a contribution area which includes ventures from the Hadoop Community. For effective booking of work, extremely Hadoop-compatible file-framework should provide area awareness: the name of the rack (more precisely, of the network switch) where a laborer hub is. Hadoop applications can use this Data to run work on the hub where the Data is, and, failing that, on the same rack/switch, lessening backbone traffic. The Hadoop Appropriated File Framework (HDFS) uses this when imitating data, to try to keep diverse copies of the Data on diverse racks. The goal is to reduce the impact of a rack power outage or switch failure so that indeed if these events occur, the Data may still be readable.

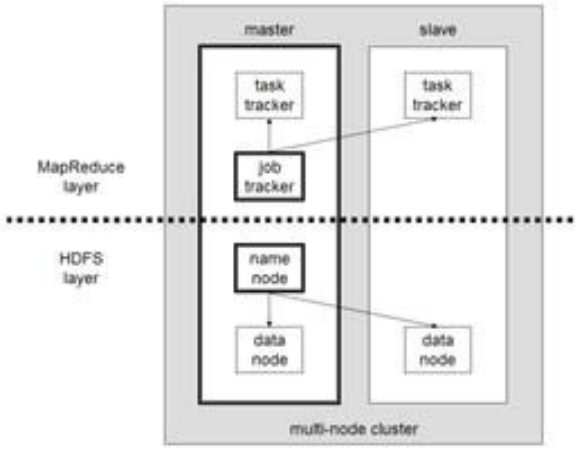

Fig 5. A Multi-Hub Hadoop cluster

As shown in Fig. 5, a small Hadoop bunch will incorporate a single master and multiple laborer nodes. The master hub comprises of a JobTracker, TaskTracker, NameNode, and DataNode.

A slave or laborer hub acts as both a DataHub and TaskTracker, though it is conceivable to have data-only laborer nodes, and compute-only laborer nodes; these are normally only utilized in non-standard applications. Hadoop requires JRE 1.6 or higher. The standard startup and shutdown scripts require Secure Shell (SSH) to be set up between hubs in the cluster.

In a greater cluster, the HDFS is managed through a dedicated NameHub server to host the fileframework index, and a auxiliary NameHub that can generate snapshots of the namenode's memory structures, thus preventing fileframework corruption and lessening loss of data. Similarly, a standalone JobTracker server can oversee job scheduling.

In groups where the Hadoop MapReduce engine is deployed against an alternate filesystem, the NameNode, auxiliary NameHub and DataHub architecture of HDFS is replaced by the filesystem-particular equivalent. One of the cost advantages of Hadoop is that since it relies in an internally redundant Data structure and is deployed on industry standard servers rather than costly specialized Data capacity systems, you can afford to store Data not beforehand viable.

Big Data is more than simply a matter of size; it is an opportunity to find insights in new and emerging sorts of Data and content, to make organizations more agile and to answer questions that were beforehand considered beyond reach. Enterprises who build their Big Data solution can afford to store literally all the Data in their organization, and keep it all online for constant interactive querying, business intelligence, investigation and visualization.

V. CONCLUSIONS

The year 2012 is the year when associations are beginning to orient themselves towards the use of Big Data. That is why this article presents the Big Data idea and the technologies associated in request to understand better the multiple benefices of this new idea ant technology.

In the future we propose for our research to further investigate the practical advantages that can be gain through Hadoop.

References:

[1] Jinson Zhang; "BigData visualization: Parallel coordinates using density approach" Published in: Systems and Informatics (ICSAI), 2014 2nd

International Conference on Date of Conference: 15-17 Nov. 2014 Page(s): 1056 – 1063.

[2] Yu, M. "Invited Speech: BigData application for 28nm and below "Published in: e-Manufacturing & Design Collaboration Symposium (eMDC), 2012 Date of Conference: 4-4 Sept. 2012] Page(s): 1

[3] D'Ambrosio, R.; Belhajali, W.; Barlaud, M. "Boosting Stochastic Newton Descent for Bigdata large scale classification" Published in: Big Data (Big Data), 2014 IEEE International Conference on Date of Conference: 27-30 Oct. 2014 Page(s): 36 - 41.

[4] Devgun, A.; "A machine learning adaptive approach to remove impurities over Bigdata" Published in: Electronics, Communication and Computational Engineering (ICECCE), 2014 International Conference on Date of Conference: 17-18 Nov. 2014 Page(s): 220 - 225

[5] Ramamoorthy, S.; "Optimized data analysis in cloud using BigData analytics techniques" Published in: Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on Date of Conference: 4-6 July 2013 Page(s): 1 – 5

[6] Shidaganti, G.; Prakash, S. "Feedback analysis of unstructured data from collaborative networking a BigData analytics approach" Published in: Circuits, Communication, Control and Computing (I4C), 2014 International Conference on Date of Conference: 21-22 Nov. 2014 Page(s): 343 - 347

[7] Pradhananga, Y.; Karande, S.; Karande, C. "CBA: Cloud-Based Bigdata Analytics" Published in: Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on Date of Conference: 26-27 Feb. 2015 Page(s): 47 – 51.

[8] Arun, J.; Hazaruthin, M.M. ; Karthik, M. "Analytics as a service delivery model for the cloud" Published in: Engineering and Technology (ICETECH), 2015 IEEE International Conference on Date of Conference: 20-20 March 2015 Page(s): 1 - 5

[9] Ghazali, N.; Shahidan, W.N.W. ; Din, Z. ; Rahim, A. more authors "Remote data acquisation and QC tools for vehicle navigation mapping solution" Published in: Space Science and Communication (IconSpace), 2011 IEEE International Conference on Date of Conference: 12-13 July 2011 Page(s): 217 – 220.

[10] Yasar, Onur; "Developing easy pluggable test equipments for factory and development testing of air defense gun system" Published in: IEEE AUTOTESTCON, 2015 Date of Conference: 2-5 Nov. 2015 Page(s): 169 – 172.

[11] Konik, A. ; Johnson, K.L.; Dasari, P. "Digital anthropomorphic phantoms of non-rigid human respiratory and voluntary body motions: A tool-set for investigating motion correction in 3D reconstruction" Published in: Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE Date of Conference: 23-29 Oct. 2011 Page(s): 3572 – 3578

[12] Yen-Feng Li; Taipei, Taiwan; Pai-Chi Li "Ultrasound beamforming with compressed data" Published in: Ultrasonics Symposium (IUS), 2010 IEEE Date of Conference: 11-14 Oct. 2010Page(s): 1912 – 1915.

[13] Xinzhong Liu; Gaochao Xu; Xiaodong Fu; Yushuang Dong "Test Data Generation Considering Data Dependence" Published in: Frontier of Computer Science and Technology (FCST), 2010 Fifth International Conference on Date of Conference: 18-22 Aug. 2010 Page(s): 208 – 213

[14] Romli, R.; Sulaiman, S.; Zamli, K.Z. "Test data generation framework for Automatic Programming Assessment" Published in: Software Engineering Conference (MySEC), 2014 8th Malaysian Date of Conference: 23-24 Sept. 2014 Page(s): 84 - 89

[15] Hameed, S.A.; Int. Islamic Univ. Malaysia, Gombak, Malaysia; Deraman, A. ; Hamdan, A.R. "Framework for Intelligent Meaningful Test Data Generation model-IMTDG" Published in: TENCON 2000. Proceedings (Volume:2) Date of Conference: 2000 Page(s): 16 - 22 vol.2

[16] Zhao, R. ; Qing Li "Automatic Test Generation for Dynamic Data Structures" Published in: Software Engineering Research, Management & Applications, 2007. SERA 2007. 5th ACIS International Conference on Date of Conference: 20-22 Aug. 2007 Page(s): 545 – 549

[17] Alnafoosi, A.B.; Coll. of Comput. & Digital Media, DePaul Univ., Chicago, IL, USA; Steinbach, T. "An integrated framework for evaluating big-data storage solutions - IDA case study" Published in: Science and Information Conference (SAI), 2013 Date of Conference: 7-9 Oct. 2013 Page(s):947 – 956

[18] Roos, A.; Drusedow, S. ; Hosseini, M.I. ; Coskun, G. "Trust Level Based Data Storage and Data Access Control in a Distributed Storage Environment" Published in: Mobile Cloud Computing, Services, and Engineering (MobileCloud), 2015 3rd IEEE International Conference on Date of Conference: March 30 2015-April 3 2015 Page(s): 169 – 176

[19] Wuling Ren ; Hangzhou; Pan Zhou "Research on the Data Storage and Access Model in Distributed Environment" Published in: Computer Engineering and Technology, 2009. ICCET '09. International Conference on (Volume:1) Date of Conference: 22-24 Jan. 2009 Page(s): 134 – 136.

[20] Coufal, H.; "Holographic data storage or reliability; a status report" Published in: Optical Data Storage Topical Meeting, 1997. ODS. Conference Digest Date of Conference: 7-9 April 1997 Page(s): 46 – 47.