

Available online at <http://www.mecspress.net/ijwmt>

Managing Big Data: A Step towards Huge Data Security

PankajDeep Kaur, Awal Adesh Monga

GNDU Regional Campus, Jalandhar

Abstract

Big data management (BDM) is the administration, management and governance of large volumes of all types of structured, semi-structured and unstructured data. Big data management is a blend of old and new best practices, teams, data types, skills, and home-grown or industrial-built functions and methods. Big Data Management is at the genesis of massive change to technology that will spur massive societal and industrial change. In this paper, different key factors kept in mind while managing big data are discussed along with the process of managing data so that organisations can get secure and accurate data. Further different developers have introduced methods have also been discussed and with it the main tools used in management of Big Data are introduced in this paper. BDM is suffering from many issues in the world are also explained here. Big Data Management deals with the integration, manipulation, quality and governance and management of Big Data including the key factors- Volume, Velocity and Variety of Big Data.

Index Terms: Hadoop, Big Data Integration, Big Data Manipulation, Big Data Quality, Big Data Governance and Management.

© 2016 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science

1. Introduction

Big data management is about two things—big data and data management—plus how the two work together to achieve business and technology goals[9].Big Data Management serves as the basic step for managing and administrating large amount of data called as Big Data in the organizations. The term “big data” has become a requirement for the companies. While meeting this need of the companies and organizations they are facing the beast of big data with the problem of managing all corporate data, and in some cases, as an all-encompassing new phase for organizational data whether structured or non-structured.

Big data technologies have been developed to handle very large volumes of dynamic and non-structured data which provide large organizations with an idealized platform for centralizing and storing their all kinds of data, yielding major advantages in scope of analytics, new opportunity identification, and knowledge discovery. This

* Corresponding author.

E-mail address: pankajdeepkaur@gmail.com, awaladesh@gmail.com

paper is aimed at technology-aware business managers, IT researchers and all those who recognize that many upcoming applications need new data-management techniques and methods.

As businesses, nonprofits and governments realize the importance of data—not only for day-to-day functions, but to their strategies and success going forward. Organizations are capturing and converting more and more content into digital data. Data types such as video, which used to be considered nice to have on a company's website, have become essential marketing, training and communications tools. Advances such as special effects and high definition significantly increase the amount of data generated; it takes twice as much space to store 3-D video as 2-D video because the technique requires two cameras to shoot the same footage, putting a significant strain on storage resources [8]. This forces developers to invent new techniques and tools so that this large amount of data can be managed and used anywhere and everywhere.

Big data technologies have already revolutionized the way Facebook, Groupon, Twitter, Zynga and countless other new business models have all been made possible because of this advance. This is a paradigm shift in technology that may end up bigger than the commercialization of the Internet in the late nineties. Entire industries and markets will be impacted on how to use these capabilities to not only provide better delivery and advanced functionality in the products and services we provide today, but we will be able to deliver new capabilities that were once only a dream[2].

The privacy of data is another huge concern, and one that increases in the context of Big Data. For electronic health records, there are strict laws governing what can and cannot be done. For other data, regulations, particularly in the US, are less forceful. However, there is great public fear regarding the inappropriate use of personal data, particularly through linking of data from multiple sources. Managing privacy is effectively both a technical and a sociological problem, which must be addressed jointly from both perspectives to realize the promise of big data [7]. Not only the storage of data is essential but the privacy and integrity of big data should also be ensured in Big Data Management. Only if the privacy of data will be maintained then only it can be assured that data is true and accurate.

Management of Big Data does not only cover the area of managing Big Data. It also helps in organizing and retrieving of large amount of data which includes all types of structured and unstructured data. Unstructured data includes all types of data that is not arranged in any sort of manner and is irregular.

Governance of Big Data is also the most important aspect under the management of Big Data. Governance not deals with the storage of data but it also deals with the different types of risks associated with the management of Big Data. Governance ensuring that corporate and governmental rules and policies are adhered to using policies, processes, and controls[1]. It helps us from the unauthorized access to the Big Data which needs full security and supervision. This feature of governance is provided by the Big Data management.

In Section I, the introduction to Big Data Management is given. Section II deals with the survey on various aspects kept in mind while managing large amount of data in all sectors of industries adopted and proposed by different researchers and developers. Different developers have proposed different approaches and tools for this purpose of management. Developers are still working and finding out different kinds of tools and methods so that management of data becomes simpler and easier. They are still working in this field because this field has no end. It becomes more and more complex when more and more amount of data is coming to be saved from various parts of life and enterprises. This covers all the key points which were undertaken by different researchers while developing different methods for managing of Big data. This mainly shows the importance of various aspects of big data.

2. Related Work

An efficient data management approach not only permits Big Data to be stored up far more easily and effectively but also makes it more recoverable and useable with a cost savings of about 90% - while decreasing the responsibility of IT staff to drive more applicable and strategic technology methods that drive growth of company instead of engaging in a silent war with an out-of-control beast of Big Data [12].

Dealing with Big Data is such a complex work such that its management is very difficult. This is mainly because there is large amount of data and also a lot kind of data which includes many types of data like audio, video, tables, documents etc. This is not only the problem. Data stored at different locations such that having different geographical locations nearly makes it impossible to manage such large amount of data. Managing Big Data plays important role here in large scale organisations.

Big Data is in various fields as recommendation engines like Amazon, Facebook, Google, used in marketing campaign analysis, Customer retention, Social Graph, Capital Markets Analysis, Predictive Analytics, Risk Management, Rogue Trading, Fraud Detection, Retail Banking, Network Monitoring, Research and Development. These all require the proper management of data. Thus they proposed different methods and tools in order to deal with the Management of Big Data.

Vinayak Borkar[3] proposed ASTERIX software using Hadoop technique and Parallel Database Management systems. ASTERIX could handle and manage all kinds of database in the form of relational database models. So the queries used for handling and managing data were very complex. Moreover the complexity of the software stack was not able to manage unstructured data. So this became the disadvantage of this software.

George Gilbert[5] in 2012 explained about the importance of 3 V's in management of Big Data and also proposed Cassandra for workload challenges in Big Data Management. But according to the increasing demand for Big Data Management it was not enough. So new technologies and techniques have been proposed for managing data in an efficient manner.

Ash Ashutosh[12] in his paper described the steps and approach to manage data. He focused mainly on the virtualization of Big Data. He is the only man who considers virtualization as the main focus in order to manage the Big Data.

Talend in 2012[4] published a white paper which explains about the Total Data Management(TDM) concerns over the ability to integrate different data such as structured and unstructured formats rapidly and easily into big data systems, including the use of new technologies was the key concern of enterprise IT staff by a clear margin. Its advantage is that it could handle all sorts of data. But the disadvantage of this is that not maximum industries have accepted the use of big data. So it becomes difficult to manage all the data.

Talend in 2013[2] published a research paper which includes the main 4 key features(big data integration, manipulation, quality, management and governance) and revolves around these features of big data. Big data in itself is very big so it is a challenge to manage the data keeping in mind these feature related to it.

Doug Cackett[10] in 2013 explained about 4V's which includes Volume, Velocity, Variety and Value in order to manage the Big Data. They used Apache's Hadoop and No-SQL databases in order to manage the Big Data. He also explained the data warehouse and the working of different technologies in the working, organizing and managing of Big Data. In this white paper we have set out Oracle's Reference Architecture for Information Management, showing how weakly typed data can be incorporated into the value creation flow and highlighted the two dominant approaches being adopted by the data driven companies.

Fern Halper[1] in 2014 focuses only on 3 V's related to big data i.e. Volume, Velocity and Variety which encompasses only the quality of data. It covers all the aspects of data but only the quality of big data is not of our concern. Big Data has to be managed in all its aspects so that it can describe the real meaning of management.

Min Chen ·Shiwen Mao ·Yunhao Liu [17] explained about the dependency of big data on cloud computing along with its relationship with the IoT (Internet of Things). They found that big data is generated by all of these technologies so they have different methods of managing data. They explained the concept of emerging value of big data.

Jinchuan CHEN, Yueguo CHEN, Xiaoyong DU, Cuiping LI, Jiaheng LU , Suyun ZHAO, Xuan ZHOU[18] clearly defines that the biggest challenge of Big Data is the management of data. He discussed various ways in which data is being managed till the date.

Hongtaek Ju, Choong Seon Hong, Makoto Takano, Jae-Hyoung Yoo, Kuang-Yao Chang, Kiyohito Yoshihara, Jeu-Yih Jeng[19] gave a report dealing with the management of Big Data with the IoT. They gave a report summarizing the relationship of IoT and Big Data

Ohbyung Kwona, Namyeon Leea, Bongsik Shinb[20] proposed a model to predict the intention to obtaining data analytics through the lens of RBV and ‘internal’ (rather than ‘external’) isomorphism.

These developers proposed different methods and different key factors in order to manage data. These techniques are still not enough to manage Big Data. So all the researches are working on proper management so that data retrieved is accurate and error free.

3. Specifications of Bdm

There are various other general things which are needed to be explored in order to understand Big Data Management more effectively. These are discussed in this section.

3.1 Factors affecting Management

There are a lot of governance and management tools to choose from [13]. These tools are Map Reducing, Hadoop, Pig, Hive, HBase, HCatalog, Flume, Oozie, Mahout, Sqoop, NoSQL[1]. These are the tools used by different organizations to store and manage the Big Data. The proper working of these tools is needed to be understood so that management of data could become easier. Thus the complexity of big data is needed to be known and then only these tools can be used.

The management part of big data may be defined in terms of the Three Vs: volume, velocity and variety which are the main characteristics of Big Data.

- Volume – the quantity of data in relation to the capability to store and manage it. Not just collecting commercial transactions and moving samples and adding to another database for further analysis, applications now collect all possible data for analysis.
- Velocity – the speed of calculation needed to get the output of the data relative to the rate of change of the data. Traditional transaction-processing applications would have collected various transactions in real time from end users, but newer technologies are increasingly collecting data coming in from other systems. Traditional applications also move their selected data to an enterprise data warehouse through an easy and attentive process that focuses on historical analysis.
- Variety – a calculation of the number of different formats of the data (e.g. text, audio, video, logs etc). The variety of data is much increasing now, because data no longer comes alone from business transactions. It mostly comes from sensors, machines and various unrefined sources, making it very difficult to manage[5].

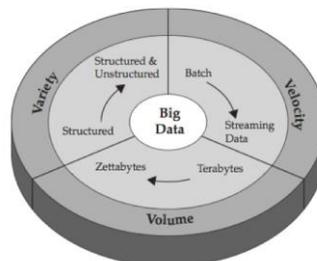


Fig 1. Three characteristics of Big Data[5]

Big data technologies that have been developed by the organizations till the date are able to handle quickly changing and non structural data with the centralized platform and thus this type of data can be used for knowledge purposes, analytics, new opportunity identification such that it provides high speed, high performance storage of data and retrieval of data which in many cases reduces the hardware costs. To address the problems of management, there are four key points to be considered when building a big data management solution:

- **Big Data Integration-** Loading big data (large amount of log files, various data from operational systems, sensors, social media, or other sources) into Hadoop via HDFS, HBase, Sqoop or Hive is considered an operational data integration problem.
- **Big Data Manipulation-** There is a range of tools that enable a user or consumer to take advantage of big data parallelization to perform transformations on large amounts of data. These languages such as Apache Pig provide a scripting language to remove errors, compare, compute and group data into an HDFS cluster.
- **Big Data Quality-** Proper management offers data quality that take the advantage of parallel computing of Hadoop. These big data quality features provide explicit functions and tasks to profile and identify duplicity of records across the huge data stores in moments not days. This is a natural extension of industrial data quality and data integration solutions and best practices.
- **Big Data Management and Governance-** While most of the early big data projects are not dependent on explicit project management structure, this needs to be changed as they become part of the bigger system. With that change, companies will need to consider standards and procedures around these projects just as they have with data management projects in the past[2].

3.2 Existing methodologies in BDM

If corporate unstructured and semi-structured data is generated largely within organizations, big data is driven by the increasing convergence of organizations with Web technologies. Although the Web is responsible for generating much of the data that most organizations would consider big data, as well as developing most of the technology to deal with it, not all big data is Web generated. Many organizations, including those in finance, insurance and Government, face significant big data processing challenges that have little or nothing to do with Web use.

The combination of massive volumes and large variety of formats that characterizes big data means that conventional databases can struggle to process information in a timely fashion. This has the drastic effect of greatly reducing the types of analysis that organizations can effectively run on very large bodies of data. Big data technologies such as Hadoop meet this challenge by dividing conventional databases – which store the data – with technologies that distribute, or map, the processing across multiple servers, and then recombine, or reduce, the result set, and return it to the database or application. This allows data mining and analytics to be performed very rapidly, while the number of factors that can typically be analyzed can often be multiplied by hundreds.

One example of the performance advantage of big data techniques over SQL relational approaches is provided by Equifax, the credit reporting agency. Another approach does away with underlying SQL databases altogether, replacing them with databases that are capable of very rapid record storage and retrieval, and greater flexibility and efficiency in handling very different types of data. Both Google and Amazon have pursued this course in their web infrastructures, while Facebook and Twitter have infrastructures based more on mapping and reducing data derived from more conventional database designs. This addresses both the second and the third of the ‘V’s: velocity and variety[4].

Another approach given by Ash Ashutosh[12] included various steps for managing the Big data. The first step for managing large amount of data is to reduce the amount of data by breaking data into unique sets.

Second step is to multiply the power of virtualization in the industries. Industries must virtualizes this data set so that not only multiple applications and users can use the same large amount of data, but also the smaller data can be saved on any independent storage device. Virtualization is the powerful weapon used for managing big data and also the other challenges of managing big data. By following these steps of management big data is changed into virtual data. Thus data management is improved in the industrial sector. Virtualization is thus the backbone when it comes to managing Big Data. And, it gives industries other benefits like flexibility, decreased costs and freedom from IT centralized data.

Big data systems also set out to preserve the variety and scope of source data. Where possible, according to big data advocates, keep everything. Converting data to be stored in a relational database often strips away vital information.

Dedicated NoSQL databases can also be designed for specific data types, such as XML, graphs, or documents. Their storage and retrieval can be orders of magnitude both faster and more efficient, and they can retain all of an object’s full information than if it is converted to fit into an SQL schema, and can be far more complex in extracting meta-data from variable or differently sourced semi-organized data such as XML or JSON than relationally structured databases, which require more static and set data model for data extraction.

4. Tools and Techniques

Different tools have been developed by different developers in order to manage huge data. Different tool have its own method of implementation and management. This functionality of tools depend on the aim for which that particular tool was developed.

Many architecture for data management has been developed till the date. The basic functionality of these tools can be divided into two types of management:

- Master-job management- which examines the processing nodes assigns tasks to the nodes and manages their activity.
- Storage management- it examines the data storage and divides the datasets to different database.

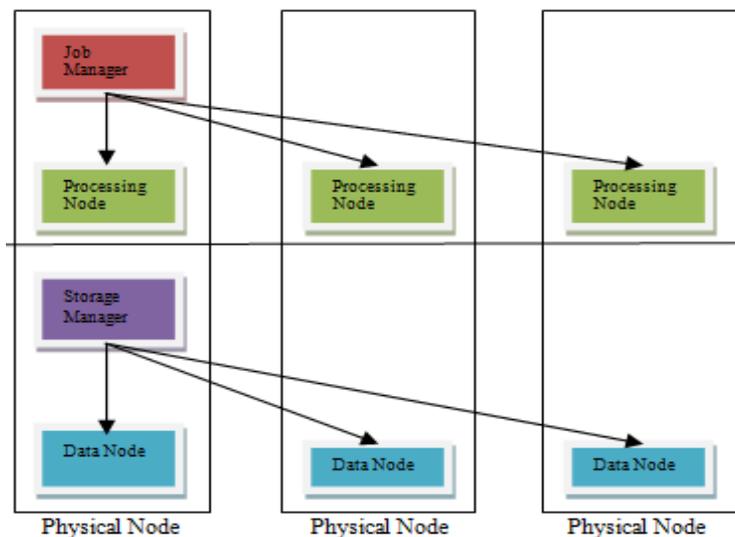


Fig 2. Big Data Management

For this there are different tools developed which are discussed below

4.1. HDFS

Hadoop Distributed File System (HDFS) distributes the data over the data nodes. There are four types of nodes involved within HDFS. They are:

- Name Node: a facilitator that provides information on the location of data. It knows which nodes are available, where in the cluster certain data resides, and which nodes have failed.
- Secondary Node: a backup to the Name Node
- Job Tracker: coordinates the processing of the data using Map Reduce.
- Slave Nodes: store data and take direction from the Job Tracker.

A Job Tracker is the entry point for a “map job” or process to be applied to the data. A map job is typically a query written in java and is the first step in the Map Reduce process. The Job Tracker asks the name node to identify and locate the necessary data to complete the job.

Once it has this information it submits the query to the relevant named nodes. Any required processing of the data occurs within each named node, which provides the massively parallel characteristic of Map Reduce.

When the each node has finished processing, it stores the results. The client then initiates a "Reduce" job. The results are then aggregated to determine the “answer” to the original query. The client then accesses these results on the file system and can use them for whatever purpose [1].

4.2. MapReduce

MapReduce is the combination of job management and the models used in programming for execution. It basically works on the principle of master-slave model in which one node acts as the master node called as the Job Tracker and the other node as slave node called as the Task Trackers.

- Job Tracker- it manages the resources acting as the slave nodes. It continuously monitors the availability as well as the accessibility of resources acting as the slave nodes so that the execution can be completed in time and efficiently. It also has the fault tolerance capability.
- Task Tracker- these are the nodes actually performing the task. The working of these nodes is simpler but these are the actual worker who performs the task. When the task or job is assigned to it, it accepts the task according to the availability and then performs the execution.

There are mainly three limitations in this model.

- Locality- the applications closer to the processing and data are easily provided with the outputs but those who are farther always suffer from the latency issues of the network.
- Mapping- applications are not easily mapped on these models. So this creates the problem of execution of the applications.
- Allocation- allocation of the processing nodes is fixed in some cases. Thus when we compare the allocation of tasks to the nodes several nodes come out to be unused. Thus the proper allocation of processing nodes is not done in this model.

These issues of MapReduce are being tackled in new model called YARN. YARN has the centralized resource management and each node have its local Node Manager which manages the task assigned to the node and the execution of the task. This allows the flexibility of assigning jobs or tasks to the nodes. Thus it reduces the latency issues of the applications and the performance of the model is improved.

4.3. Zookeeper

When there are multiple jobs being performed in single distributed system then the developer needs to manage the working of the system and have to maintain the synchronization between the various nodes of the same system. Zookeeper is a centralized service for maintaining configuration information, providing group services, naming and providing distributed synchronization.

Thus Zookeeper is used to manage the process of naming the objects and then further manages these objects in the hierarchical manner. Due to synchronisation capability of Zookeeper it is also used to control the shared resources such that no deadlock state is met.

4.4. HBase

HBase is the non relational data management. It does not support SQL queries. It divides the huge datasets such that these datasets could be used to acquire the required data. The basic development of HBase was from Google's Big Table. So HBase is column oriented. HBase supports memory execution because the column oriented data can be compressed such that more amounts of data can be represented.

The main advantage HBase is that it acts as an alternative for data storage in MapReduce applications. Because of the columnar orientation and the data organizing capacity of HBase.

4.5. Hive

Although MapReduce have methodologies for executing different applications and develop them but companies or organizations relies on the tabular representation of data. Thus this problem of MapReduce gave basis to the development of Hive which comes under the category of Data Warehousing.

Hive enables the data organization in data warehousing. Hive works on the query language called HiveQL which is same the SQL. This system provides the way of loading, extracting and transforming data. It also acts as the native access to MapReduce programmers because the users or developers or programmers can integrate the Map and Reduce Functions to HiveQL queries such that it can programs can be used in Hive also. Thus it reduces the effort of programmers to write program again and again.

4.6. Pig

It is based on the process of abstracting some details by using a higher level programming language called Pig Latin. It allows the user to specify the process of analysis is being performed. It works in similar way as the SQL works on structured dataset. The similarity exists when removing the duplicate data, splitting datasets, splitting datasets, joins etc. It also allows user to create new user defined functions which can be further used in other programs.

5. Issues in Bdm

Big data management is an issue in itself but today different tools have been developed for managing this huge amount of data. There are various issues related to BDM. Not everyone has the same issues. Issues differ from organisation to organisation. Still these issues overlap. But mainly there are three issues which are:

- Skill – skill issues can be seen from two different angles. Firstly there comes the problem of organisations already have the skills but they are not able to reuse it. Thus these companies loose the efficiency of skill. Then come the problem of organisations not having the skills. These types of organisations need to acquire these skills that will take time and also money.
- Integration- the process of integration is different for different companies. So this creates confusion on how to integrate and also how to query the distributed large data.
- Security- security is not provided by the tools and applications to its fullest. This field needs to be worked more. Due to the large amount of data there are more chances of error and unauthorised access to data. These cause security issues while managing big data.

Storage- storage of huge data is an important issue for the management of big data. Management is possible only and only when the storage of our data is done properly and efficiently such that retrieval of data from huge data sets is simpler and user friendly.

6. Conclusions

Big data shows an essential paradigm shift in industrial technology enabling companies to gain a step into new business opportunities (and problems). Big data management is the administration, management and governance of large volumes of all types of data [11]. Big Data Management is the most common problem faced by almost all the organizations which have accepted Big Data as the best approach to store all types of data whether structured or unstructured. Management of Big Data is one of the fields which are still under work because its complexity is increasing day by day as the volume of Big Data is increasing at an alarming rate. Management of Big Data includes the governance and managing of data such that the quality of data is not compromised. Thus this forms the basis of Big Data Management including the Big Data Integration and Big Data Manipulation. Management does not end with this only. While management of Big Data the main characteristics of data are also kept in mind-Volume, Velocity and Variety. Hadoop forms as the basic tool which is used till the date and will be used in the coming years also in combination with the different tools as per the requirement of the organizations. Different tools are being used like MapReduce, Zookeepers, Hive, HBase and Pig which are accepted by the organizations in order to solve their problem of managing data. Management of Big Data is important because in today's world almost all the organizations have a large amount of data which needs to be managed and controlled so that the data of these does not crash. Thus smooth and proper functioning of Big Data requires the management of Big Data. Big Data Management is also facing many issues due to the difference in the objectives of the organizations. Different organizations have adopted different methods and tools according to their requirement to manage their data and thus have proposed new techniques for management. Thus management deals with all the aspects of Big Data so that the privacy, authenticity, integrity of data is retained.

Acknowledgement

This paper is not the initiative of mine alone. There are many persons attached to me and my paper. I would like to thank my professor and guide **Pankajdeep Kaur** who gave this topic for the research. She showed me the path to complete my research. I would like to thank God who gave me the patience to do this job and helped me in completing my task. I would like to thank my family who supported me in my work.

References

- [1] Fern Halper, "Big Data Management", TDWI E-Book, January 2014.
- [2] Talend White Paper, "Four Key Pillars for Big Data Management Solution", 2013.
- [3] Vinayak Borkar, "Inside "Big Data Management":Ogres, Onions, or Parfaits?", EDBT Association, 2012.
- [4] Talend White Paper, "A Total Data Management for Big Data", September 2012.
- [5] George Gilbert, "A Guide To Big Data Workload Management Challenges", *GigaOM PRO*, May 2012.
- [6] www.michael-noll.com/tutorials/running-Hadoop-on-ubuntu-linux-single-node-cluster/.
- [7] Philip Bernstein, "Big Data White Paper", Microsoft, Feb 2012.
- [8] Eric Bassier, "Big Data: Managing Explosive Growth", *Quantum*, Jan 2012.
- [9] Philips Rossum, "Managing Big Data", TDWi Research, fourth quarter 2013.
- [10] Doug Cackett, "Information Management And Big Data A Reference Architecture", Oracle, February 2013.
- [11] <http://searchdatamanagement.techtarget.com/definition/big-data-management>.
- [12] Ash Ashutosh, "Best Practices For Managing Big Data", Forbes, May 2012.
- [13] Michael Biddick, "The Big Data Management Challenge", InformationWeeks, April 2012.
- [14] https://blogs.oracle.com/bigdata/entry/3_key_problems_to_solve.
- [15] <https://www.safaribooksonline.com/library/view/big-data-analytics/9780124173194/xhtml/CHP007.html>.
- [16] <http://www.computerweekly.com/opinion/How-to-manage-big-data-and-reap-the-benefits>.
- [17] Min Chen · Shiwen Mao · Yunhao Liu, "Big Data: A Survey", Springer Science+Business Media New York 2014, 22 January 2014.
- [18] Jinchuan CHEN, Yueguo CHEN, Xiaoyong DU, Cuiping LI, Jiaheng LU , Suyun ZHAO, Xuan ZHOU, "Big data challenge: a data management perspective", Higher Education Press and Springer-Verlag Berlin Heidelberg 2013.
- [19] Hongtaek Ju , Choong Seon Hong, Makoto Takano, Jae-Hyoung Yoo, Kuang-Yao Chang, Kiyohito Yoshihara, Jeu-Yih Jeng, "Management in the Big Data & IoT Era: A Report: on APNOMS 2012", Springer Science+Business Media New York 2013, 22 february 2013.
- [20] Ohbyung Kwona, Namyoon Leea, Bongsik Shinb proposed, "Data quality management, data usage experience and acquisitionintention of big data analytics", International Journal of Information Management 34 (2014), 13 march, 2014.

Authors' Profiles



Dr. Pankaj Deep Kaur is working as an Assistant Professor in the Department of Computer Science and Engineering, Guru Nanak Dev University, RC, Jalandhar, India. She received her Bachelor's Degree in Computer Applications (2000) and Master's Degree in Information Technology (2003) from Guru Nanak Dev University, Amritsar, India. She completed her Ph.D. in Resource Scheduling in Cloud Computing from Thapar University, Patiala (2014) and has over ten years of teaching and research experience. She has been a university position holder in her graduation studies and received Gold Medal for her excellent performance in her Post Graduation studies. She is a recipient of Junior Research Fellowship from Ministry of human Resource and Development, Govt. of India. Her research interests include Cloud Computing and Big data.



Awal Adesh Monga is currently pursuing her post graduation from Guru Nanak Dev University Regional Campus Jalandhar. She received her Bachelor's Degree in Computer Science and Engineering (2014) from Malout Institute of Management and Information Technology (MIMIT), Malout, Punjab.

How to cite this paper: PankajDeep Kaur, Awal Adesh Monga, "Managing Big Data: A Step towards Huge Data Security", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.6, No.2, pp.10-20, 2016.DOI: 10.5815/ijwmt.2016.02.02