Computation of Big Data in Cloud Environment using Hadoop

Saurabh Sindhu
Assistant Professor, CRM Jat College, Hisar – 125001, Haryana

Divya Sindhu
Assistant Professor, CRM Jat College, Hisar – 125001, Haryana

Abstract - Big data is a collection of large data sets that exceeds the processing capacity of conventional database systems. Since the rise of digitilization, organizations from various sectors have amassed huge amounts of digital data, capturing trillions of bytes of information, ranging from the customers, to suppliers and operations. Various technologies like Hadoop, MapReduce and NoSQL etc. are used to process Big Data, but these technologies are very costly and cannot be afforded by the mid- and small-size organizations. Therefore, there is a need of cost-effective technology, which could be afforded by everyone. Several systems have been developed including cloud computing that support Big Data. The cloud provides a reliable, accessible and scalable environment for Big Data systems to function. It provides the users with low maintenance cost and broad network access. The hosting companies, which have advanced data centers, rent spaces that are stored in a cloud to their customers in line with their needs. Thus, cloud computing and Hadoop has become a new distributed storage model for most of the organizations and industries. However, relying on a single cloud storage provider has resulted in problems like service level agreements (SLAs) and vendor lock in. Therefore, multi cloud environment is used to address the security and the data availability problems. In this paper, various Big Data processing methods, approaches, challenges and implementation of cloud computing are described. A system has been proposed that uses Hadoop computing platform in multi-cloud domain for storing customer's data reliably.

Keywords - Big Data, Hadoop, Cloud computing, Customers, Security

I. INTRODUCTION

Big Data is collection/combinations of large data sets whose size (volume), complexity (variability) and rate of growth (velocity) make them difficult to be managed, processed or analyzed by conventional computing technologies and tools [1, 2]. The data does not fit the structures of existing database architectures. The digital data is produced from different sources of technology; for example, sensors, digitizers, scanners, numerical modeling, mobile phones, internet, videos, e-mails and social networks. The data types include texts, geometries, images, videos, sounds and combinations of each. Volume of data is also growing exponentially due to machine-generated data (data records, web-log files and sensor data) and from growing human engagement within the social networks. Variety of data has also increased from text to audio, images and videos. Smartphones, E-commerce and social networking websites are examples where massive amounts of data are being generated. Big Data applications are almost ubiquitous; from marketing to scientific research to customer interests and so on. Big Data can be witnessed in action from Facebook, which handles over 40 billion photos from its user base to CERN's Large Hydron Collider (LHC) which generates 15PB a year to Walmart, which handles more than 1 billion customer transactions in an hour. Every day, we create 2.5 quintillion bytes of data, so much that 90% of the data in the world today has been created in the last two years alone. The velocity at with this data is growing is tremendous and the growth of data cannot be restricted.

Big Data can be decomposed into three layers, including Infrastructure Layer, Computing Layer and Application Layer (Fig. 1). This data can be used for benefit of public by extracting the hidden wealth of information inside it. New technologies are required to manage this Big Data phenomenon [3]. Big Data technologies include new generation technologies and architectures which are designed so as to extract value economically from very huge volumes of heterogeneous data. This is done by enabling high velocity capture, discovery and analysis. Various challenges are associated with the use of Big Data, which include scale, heterogeneity of data, lack of structure, error-handling, privacy, timeliness, provenance and visualization, data acquisition, result interpretation, data management and security [4, 5]. The Big Data generated from the large companies like Facebook, Yahoo, Google, YouTube etc. need analysis of this enormous amount of data to extract the large amount of information. There are various technologies in the market from different vendors including Amazon, IBM, Microsoft etc. to handle Big Data.

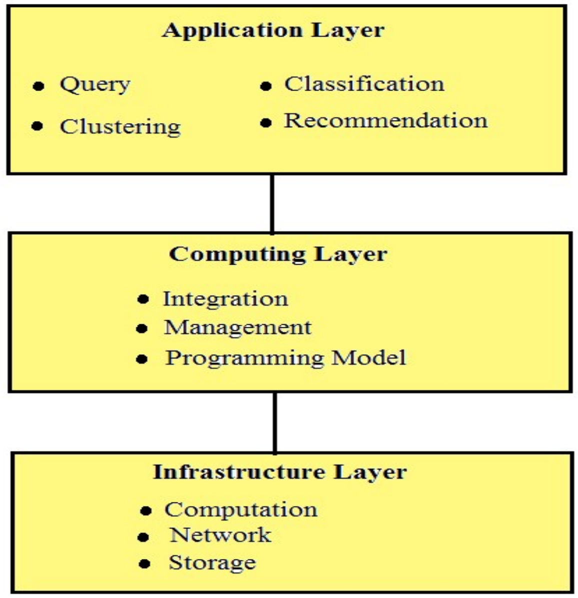

Figure 1. Three Layers of Big Data

Cloud computing refers to on-demand computer resources and systems available across the network that can provide a number of integrated computing services without local resources to facilitate user access [6]. Cloud provides the services with the ability to ingest, store and analyse data that are available to handle most of the challenges posed by Big Data. Early adopters of Big Data on the cloud are probably the users deploying Hadoop clusters on the highly scalable and elastic environments provided by Infrastructure-as-a-Service (IaaS) provider such as Amazon Web Services, for test and development, and analysis of available datasets. It offered data storage and data back-up in a cost-effective and reliable environment that gives organizations the computing resources, which they need to store their data in both the forms i.e. structured and unstructured. On the other hand, embedded analytics engines helps to analyze the data stored on the cloud at Software-as-a-Service (SaaS) level. The analytical output is then provided to the end users through a graphical interface. The development of queries and integration to the data source on the cloud are prerequisites that organizations need to perform before the usability can be delivered. Hence cloud computing is able to provide the support for Big Data deployment. But, when it comes to the transfer of data to and from these two environments, privacy, data management, retrieval, access speed, security and integration issues can become major issues for the use of Big Data on the cloud.

II. DATA AND ITS CATEGORIES

Data may exist in a variety of forms, as numbers, text, bits and bytes or as facts that are stored in a person's mind. Data is often used so as to distinguish binary machine-readable information to textual human-readable information [7]. In today’s world, the information and patterns help us in making decisions. To achieve the information and knowledge from Big Data, many methods and concepts were adopted and widely practiced is Data-ware housing [8]. There are three categories of data.

2.1. Structured Data

Data which resides in a fixed field within a record or file is called structured data. Data contains relational databases and spread sheets. Data type includes numeric, currency, alphabetic, name, date, address and with any restrictions on the data input (number of characters, restricted to certain terms. The structured data could be easily entered, stored, queried and analyzed. Certain conditions such as high cost and performance limitations of storage, memory and processing and use of relational databases were the only solution to effectively manage the data.

2.2. Semi-structured Data

Semi-structured data does not conform with the formal structure of data models associated with relational databases, but contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Semi-structured data does have some organizational properties that make it easier to analyze. In semi-structured data, the entities belonging to the same class may have different attributes even though they are grouped together and the attributes' order is not important. Semi-structured data are increasingly occurring since the advent of the internet where full-text documents and databases are not the only forms of data anymore. Semi-structured data is often found in object-oriented databases. For example, NoSQL databases are considered as semi-structured. XML, other markup languages, e-mail and EDI are also forms of semi-structured data. XML has been popularized by web services that are developed utilizing SOAP principles. JSON (JavaScript Object Notation), is an open standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. It is used primarily to transmit data between a server and web application, as an alternative to XML. JSON has been popularized by web services developed utilizing REST principles.

2.3. Unstructured Data

Unstructured data files usually consists of text and multimedia content e.g. e-mail messages, videos, photos, audio files, presentations, webpages and various other kinds of business documents. These types of files may have an internal structure, yet they are considered "unstructured" because the data contained by them doesn't fit neatly in a database. It is estimated that 80 to 90% of the data in an organization is unstructured type. However, the amount of unstructured data in organizations is growing significantly and drastically, usually many times faster than structured databases. Some examples of machine-generated unstructured data include satellite images, scientific data, photographs and video, and radar or sonar data. Examples of human-generated unstructured data are text internal to the particular company, social media data, mobile data and website content.

III. BIG DATA PROCESSING USING HADOOP

Hadoop is a programming framework used to support the processing of large data sets in a distributed computing environment. Hadoop was developed by Google’s MapReduce that is a software framework, where an application break down into various parts. The current Appache Hadoop ecosystem consists of the Hadoop Kernel, MapReduce, Hadoop distributed file system (HDFS) and numbers of various components like Apache Hive, Base and Zookeeper. HDFS and MapReduce.

3.1. HDFS architecture

Hadoop Distributed File System is a file fault-tolerant storage system designed for storing very large files with streaming data access pattern and running clusters on commodity hardware. It is able to store huge amounts of information, scale up incrementally and survive the failure of significant parts of the storage infrastructure without losing data [9]. Hadoop creates clusters of machines and coordinates work among them. Clusters can be built with inexpensive computers. If one fails, Hadoop continues to operate the cluster without losing data or interrupting work, by shifting work to the remaining machines in the cluster. HDFS manages storage on the cluster by breaking incoming files into pieces, called “blocks,” and storing each of the blocks redundantly across the pool of servers. In the common case, HDFS stores three complete copies of each file by copying each piece to three different servers [10].

HDFS architecture is broadly divided into following three nodes, which are Name Node, Data Node and HDFS Clients/Edge Node (Fig. 2). Name Node is centrally placed node, which contains information about Hadoop file system. The main task of name node is that it records all the metadata and attributes, and specific locations of files and data blocks in the data nodes. Name node acts as the master node as it stores all the information about the system and provides information which is newly added, modified and removed from data nodes. Another Data Node works as slave node. Hadoop environment may contain more than one data nodes based on capacity and performance. A data node performs two main tasks storing a block in HDFS and acts as the platform for running jobs. The HDFS Clients/Edge Node acts as linker between name node and data nodes. In Hadoop cluster, there is only one client but there are also many clients depending upon the performance needs.

3.2. MapReduce architecture

MapReduce is defined as a programming model for processing and generating large sets of data and MapReduce framework is the processing pillar in the Hadoop ecosystem [11]. MapReduce is a data processing algorithm which uses a parallel programming implementation. There are two phases in MapReduce, i.e., the “Map” phase and the “Reduce” phase. The system splits the input data into multiple chunks, each of which is assigned a map task that can process the data in parallel [12]. Each map task reads the input as a set of (key, value) pairs and produces a transformed set of (key, value) pairs as the output. The framework shuffles and sorts outputs of the map tasks and sends the intermediate (key, value) pairs to reduce task, which groups them into final results. In a traditional data ware-housing scenario, this might entail applying an ETL operation on the data to produce something usable by the analyst. In Hadoop, these kinds of operations are written as MapReduce jobs in Java. There are a number of higher level languages like Hive and Pig that make writing these programs easier. The outputs of these jobs can be written back to either HDFS or placed in a traditional data warehouse.

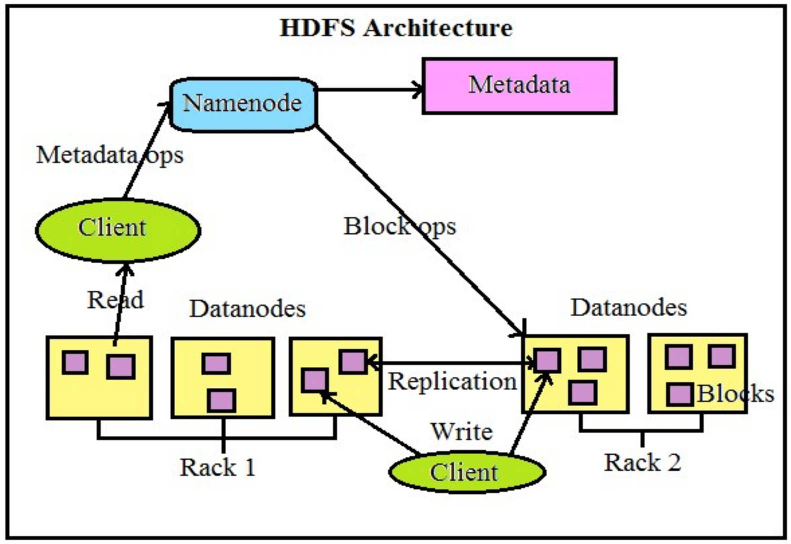

Figure 2. Hadoop Distributed File System architecture

In MapReduce paradigm, a query is made and data are mapped to find key values, which are considered to relate to the query; the results after that are reduced to a dataset which provides answer to the query. Most enterprise data management tools are designed to execute basic queries run quickly. Data is indexed so that only very small portions of the data need to be examined in order to answer a query. This solution is not applicable to semi-structured form (text files) or unstructured form (media files). To reply and execute a query successfully, all the data has to be examined [13].

The Hadoop framework utilizes MapReduce, in conjunction with additional software components like R language and a wide range of open source Not Only Structured Query Language (NoSQL) tools, which include Cassandra and Apache HBase that have become the basics of many Big Data discussions. Vendors have now launched their own versions of such tools (e.g., Oracle's version of the NoSQL database) or to integrate these tools with their own products e.g., EMC's Greenplum Community edition, which includes the open source Apache Hive, HBase and ZooKeeper [14].

IV. APPROACHES FOR BIGDATA HANDLING

NoSQL database management systems (DBMSs) are available as open source software that are designed for use in high data volume applications in the clustered environments. Because they don't adhere to a fixed schema, NoSQL DBMS permit us with more flexible usage, allowing high-speed access to both semi-structured and unstructured data. However, SQL interfaces are also being used alongside the MapReduce programming paradigm. There are several types of NoSQL DBMS like:

(i) Key-Value Stores

Key-value pair (KVP) tables are used to provide determination management for many of the other NoSQL technologies. The table has two columns; one is for the key; the other is for the value. This value could be a single value or a data block containing more than one value, the format of value(s) is determined by program code. KVP tables may use one among these: indexing, hash tables or sparse arrays to provide rapid retrieval and insertion capability. KVP tables are best for application in simple data structures and on the Hadoop MapReduce environment. Some Examples of key-value data stores include Amazon's Dynamo and Oracle's BerkeleyDB [15].

(ii) Document Oriented Database

A document-oriented database is designed for storing, retrieving and managing document-oriented or semi-structured data. The central concept of a document-oriented database is the idea of a "document" where the contents inside the document are encapsulated in some standard format such as JavaScript Object Notation (JSON), Binary JavaScript Object Notation (BSON) or XML. Examples of these databases include Apache's Couch DB and 10gen's Mongo DB.

(iii) BigTable Database

It is a distributed storage system based on the registered Google File System for managing structured data that is aimed to scale to a very large size i.e., Petabyte's of data across thousands of servers. It is also known as Distributed Peer Data Store. The volume of data handled is very high and the schema does not order the same set of columns for all rows. Each cell has a time stamp and there can be multiple forms of a cell with different time stamps. In order to manage these huge tables, Bigtable breaches tables at row boundaries and saves them as tablets. Each tablet is nearly of 200 MB in size and each server can save about 100 tablets. This arrangement allows tablets from a single table to be spread among various machines, load balancing and fault tolerance is also managed by it. An example of a BigTable database is CassandraDB [16].

(iv) Graph Database

A database that contains nodes, edges and properties to represent and store data is said to be Graph database. In this database, every entity contains a direct pointer to its neighbouring element and no index look-ups are required. A graph database is valuable when large-scale multi-level relationship traversals are public and is best suited for processing complex many-to-many connections like social networks. A graph can be taken by a table store that supports recursive joins such as BigTable and Cassandra. Examples of graph databases include InfiniteGraph from Objectivity and the Neo4j an open source graph database [17].

To opt for such Big Data processing technologies for small and medium sized industries is very expensive. Besides this Big data possess various other problems in the field of data processing and resource management, data integration, data storage, data visualization and user interaction, model building and scoring. The organized approach toward data collection in order to enhance randomness in data sampling and reduce favouritism is not apparent in the collection of Big Data sets [18]. The data collected can still be imperfect and inaccurate which, in turn, can lead to twisted conclusions. Twitter is commonly inspected for insights about user feelings. There is a natural problem with using Twitter because as a data source, only 40% of Twitter's active users are just listening and not contributing. This can suggest that the tweets are coming from a certain more vocal type of people, who are more participative in social media than from a true haphazard sample [19].Twitter makes a sample of its ingredients available to the public by its streaming Application Programming Interfaces (APIs) [20]. In broader terms, there are three areas of problems associated with Big Data which include Big Data computation and management, Big Data computation and analysis and Big Data security. The Solution that is reliable and cheap in solving the case of Big Data to benefit every class is Cloud computing [21].

V. CLOUD COMPUTING

Cloud Computing consists of network based computing which takes place over the internet or a model that depends on a large, centralized data center to store and process a wealth of information. Cloud computing is all intended to access huge amounts of computing power by combining resources to process a huge quantity of data [22, 23]. Cloud computing has become a powerful architecture to perform extensive and complex computing, and has transformed the way that computing setup is abstracted and used. An important goal of these technologies is to deliver computing as a solution for tackling big data, such as large scale, multi-media and high dimensional data sets. Cloud deployment solutions provide services that businesses would otherwise not be able to afford under the traditional hardware and software acquisition method [24]. This typical organization models for cloud computing includes: infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS) and hardware as a service (HaaS) (Fig. 3).

5.1. Infrastructure as a Service (IaaS)

In Infrastructure as a Service (IaaS) model, consumers are given full liberty to manage their data on the server. Here the service provider has to be responsible for raw storage, computing power, networks, firewalls and load balancers, and this is often demonstrated as a virtual machine [25]. A client business pays on a per-use basis whenever the equipment is used to support computing operations such as: storage, hardware, servers, and networking equipment. Infrastructure as a service in cloud computing model has received most attention from the market and about 25% of enterprises planning to adopt a service provider for IaaS [26]. Some of the services available to businesses through the IaaS model include disaster recovery, computing as a service, storage as a service, data center as a service, virtual desktop infrastructure and cloud bursting, which is providing peak load capacity for variable processes.

Profits of IaaS include increased financial elasticity, choice of services, business agility, cost-effective scalability and increased security.

5.2. Platform as a Service (PaaS)

In the PaaS model, customers are provided with an operating system, programming language execution environment, database and web server. They are not concern with the either cost or management in the hardware and software layers [27]. Generally, PaaS solutions include application design and development tools, application testing, versioning, integration, deployment and hosting, state management and other related development tools [28]. Businesses attain cost savings using PaaS through standardization and high utilization of the cloud-based platform across a number of applications [29]. Some other advantages of using PaaS include lowering risks by using pretested technologies, promoting shared services, improving software security, and lowering skill requirements needed for new systems development. PaaS provides companies a platform for developing and using custom applications needed to analyse huge quantity of unstructured data at a low cost and low risk in an environment that is secure.

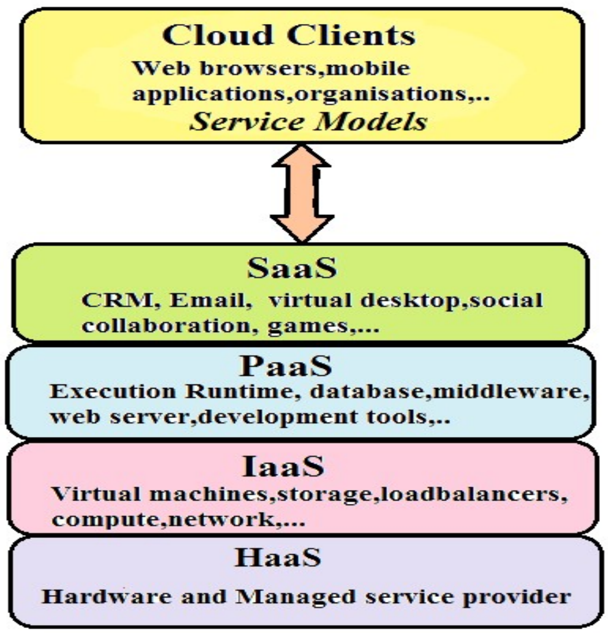

Figure 3. The different layers of cloud computing. Cloud service provides software as a service, platform as a service, infrastructure as a service and hardware as a service.

5.3. Software as a Service

In Software as a service (SaaS), consumers are given access only to the application software, which can be run remotely from the data centres of the cloud service provider [30]. Here the provider is responsible for the maintenance and support of the infrastructure and operating platforms i.e., it provides businesses with applications that are stored and run on virtual servers in the cloud. Because, cloud service providers are specialized in one area, they can provide consistent service at a fraction of the cost. The business is not charged for hardware, only for the bandwidth for the time and number of user's necessary. SaaS allows businesses to shift the risks associated with software acquisition while moving IT from being reactive to proactive. Benefits of using SaaS include easier software administration, automatic updates and patches management, software compatibility across the business, easier collaboration, and global accessibility [31].

5.4. Hardware as a Service (HaaS)

HaaS is a cloud service based upon time sharing model on minicomputers and mainframes that provide applications and services to the customers, HaaS offers only the hardware [32, 33]. In a managed service situation, the managed service provider (MSP) remotely monitors and administers hardware located at a client's site as contracted. The HaaS model allows the customer to license the hardware directly from the service provider which alleviates the associated costs. Vendors in the HaaS arena include Google with its Chromebooks for Business, CharTec and Equus.

Traditionally companies have learned how to outsource certain elements to modernise their processes. Cloud computing is the next step that allows outsourcing of IT, instead of forming and maintaining their own IT department with physical servers and technical consultants. Companies hire a cloud service company to provide all its IT needs. However, with the arrival of cloud computing, a company can acquire its exact computing needs. Thus, for companies and governments to use Big Data, cloud computing is the best option they can opt for. The cloud computing when used for Big Data computation has certain benefits like flexibility, storage, saving time and reducing costs.

VI. CLOUD COMPUTING IN RELATION TO BIG DATA COMPUTATION

It is a very tedious and complex task to implement cloud computing for Big Data processing and various other challenges associated with it. Various methods and approaches are being used for carrying out analytics on cloud for Big Data applications. Some of the areas of work include: (i) Data management (ii) Model development and scoring (iii) Visualization and user interactions, and (iv) Business model [34]. Present DBMS technologies has various issues like scalability, reliability and heterogeneity of Big Data i.e., both structured and un-structured. Furthermore, key Value stores for supporting riches set of applications is to be researched upon. This is due to the fact that cloud computing needs a DBMS that supports descriptive and deep analytics. It needs a support for updating of heavy applications, ad-hoc analytics and decision support [35]. One of the main problem rather a challenge is how security can be managed while uploading and accessing and processing of data on cloud.

Since data is heterogeneous in nature so Hadoop which works well for short jobs on homogeneous data doesn't work well. To achieve the higher rates for data retrieval, web servers can be embedded into cloud environment. This can be done by creating a multilevel index in web server with multilevel index key in data node [36]. Network intrusion is a major problem associated with cloud computing on Big Data computation. Solution to this can be integration of modern technologies, Hadoop File System and cloud technologies with latest representation learning techniques and support to predict network intrusions through Big Data classification strategies [37]. Wealth of Information needs to be extracted out of Big Data which if otherwise left untreated would make it a data monster and to overcome this we need to change our traditional mining algorithms and techniques [38]. The effective techniques for dealing with the elasticity of cloud infrastructures, designing scalable, elastic and autonomic multitenant database systems with security and privacy needs to be designed [39].

VII. CONCLUSION

Cloud computing is one of the methods that can help researchers and industry people to achieve great heights when applied on Big Data to unveil the wealth of knowledge. This information can prove boon to almost every aspect of life from agriculture to space missions. Still we need something that is effective and help us to maintain large amount of data, and also provide the security and better access, so that mining becomes an easy task to do without changing much in its existing technologies and algorithms. A perfect solution to handle heterogeneous data with security and better access on clouds is yet needed and a single perfect data management solution needs to be formulated.

REFERENCES

[1] R. L. Villars, C. W. Olofson, M. Eastwood, "Big Data: What it is and why you should care". White Paper, IDC, June 2011.

[2] S. Ikhlaq, B. Keswani, "Computation of Big Data in Hadoop and cloud environment". IOSR Journal of Engineering, Vol. 6, Issue 1, pp. 31-39, 2016.

[3] S. Kaisler, F. Armour, J. A. Espinosa, W., "Money, Big Data: issues and challenges moving forward". International Conference on System Sciences. Hawaii: IEEE Computer Society, pp. 995-1004, 2013.

[4] V.B. Bobade, "Survey paper on Big Data and Hadoop". International Research Journal of Engineering and Technology, Vol. 3, Issue 1, pp. 861-863, 2016.

[5] A. Katal, M. Wazid, R. H. Goudar, "Big Data: Issues, challenges, tools and good practices". IEEE, pp. 404-409, 2013.

[6] D. Sindhu and S. Sindhu, "Data Security and future perspectives", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 3, pp. 5840-5850, 2017.

[7] M.S. Chen, J. Han, and P.S. Yu, "Data Mining: An Overview from a Database Perspective", IEEE Transactions on Knowledge and Data Engineering, Vol. 8, Issue 6, pp. 866-883, 1996.

[8] W.H. Inmon, "Building the Data Warehouse", John Wiley. pp. 33, 1996.

[9] S. Sindhu and D. Sindhu, "Big Data Processing using Hadoop: current status and future perspectives", International Journal Computer Engineering and applications, Vol. XII, Issue IV, 2018.

[10] W.O. Carl, V. Dan, "Worldwide Hadoop – MapReduce ecosystem software 2012- 2016", Accessed at: http://www.idc.com/getdoc.jsp?containerId=234294.

[11] J. Dean, S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters". In Proceedings of OSDI '04: 6th Symposium on Operating System Design and Implementation, San Francisco, CA, USA, 2004.
[12] S. Sindhu and D. Sindhu, "MapReduce: An Efficient and Simplified Technique for Big Data Processing", International Journal of Engineering Technology Science and Research, Vol. 5, Issue 3, pp. 419-425, 2018.
[13] M. Zaharia, A. Konwinski, A. D. Joseph, R. Katz, I. Stoica, "Improving MapReduce Performance In Heterogeneous Environments", Proceedings of the 8th USENIX conference on Operating systems design and implementation, p.29-42, December 08-10, 2008, San Diego, California ,USA
[14] H. Dong, "Oracle Releases NoSQL Database, Advances Big Data Plans". Accessed at:http://www.informationweek.com/software/informationmanagement/oracle-releases-nosql-database-advances/231901480, 2012.
[15] W.O. Carl, "*The Big Deal about Big Data*". Accessed: http://www.idc.com/getdoc.jsp?containerId=226904\
[16] Google Inc., "*Bigtable: A Distributed Storage System for Structured Data*". Accessed:http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//archive/bigtable-osdi06.pdf
[17] InfoGrid."*OperationsonalGraphDatabase*",Accessed:http://infogrid.org/blog/2010/03/operations-on-a-graph-database-part-4/
[18] "*Big Data*".Accessed:http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431
[19] M. Rosoff. "*Twitter Has 100 Million And 40% Are JustWatching*", Accessed: http://articles.businessinsider.com/2011-09-08/tech/30127585_1_ceo-dick-costolo-twitter-users
[20] Twitter,Accessed:https://dev.twitter.com/docs/streaming-apis/streams/public Armbrust, Griffith, Joseph, Katz, Konwinski, Lee, Zaharia, "A View of Cloud Computing", Communications of the ACM, Vol. 53, Issue 4, 2010.
[21] Ji, Y. Li, W. Qiu, U. Awada, K. Li , "Big Data Processing in Cloud Computing Environments", International Symposium on Pervasive Systems, Algorithms and Networks, pp.17, 2012.
[22] D. Sindhu and S. Sindhu, "Cloud Computing: Data Security using Cryptographic Algorithms" International Journal of Innovations and Advancements in Computer Science, Vol. 7, Issue I, pp. 290-301, 2018.
[23] Agrawal, S. Das, and A. E. Abbadi, "Big Data and Cloud Computing: Current State and Future Opportunities", Proceedings of the 14th International Conference on Extending Database Technology, EDBT/ICDT '11, ACM. II.2(b), New York, NY, USA, pp. 530–533, 2011.
[24] O.A. Iyanda, "Big Data and Current Cloud Computing Issues and Challenges", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 6, 2014.
[25] Rouse, (2010),"*Infrastructure as a Service*", Accessed : http://searchcloudcomputing.techtarget.com/ definition /Infrastructure-as-a -Service-IaaS
[26] Cisco,(2009),"*Infrastructure as a Service: Accelerating time to profitable new revenue streams*".Accessed:http://www.cisco.com/en/US/solutions/collateral/ns341/ns991/ns995 /IaaS_BDM_WP.pdf
[27] Salesforce.com,(2012),"*The end of software: Building and running applications in the cloud*", Accessed: http://www.salesforce.com/paas/
[28] Géczy, Izumi, Hasida, "Cloud sourcing: Managing Cloud Adoption", Global Journal of Business Research, Vol. 6, Issue 2, 2012.
[29]Oracle,(2012),"*OraclePlatformAsAService*".Accessed:http://www.oracle.com/us/technologies/cloud/oracle-platform-as-aservice408171.html
[30] Carraro,Chong,(2006),"*Softwareasaservice:Anenterpriseperspective*",Accessed:http://msdn.microsoft.co m/enus/library/aa905332.aspx #enterprisertw_topic3
[31] Rouse,(2010),"*Software As A Service*", Accessed: http://searchcloudcomputing.techtarget.com /definition/Software-as-a-Service
[32] CompututureWeekly.com,(2009),"*HardwareAsAService*",Accessed:http://www.computerweekly.com/featu re/Hardware-as-a-Service
[33] Rouse,(2007),"*HardwareAsAService*".Accessed:http://searchitchannel.techtarget.com/definition/Hardwar e-as-a-Service-in-managed-services
[34] M. D. Assunção , R. N. Calheiros, S. Bianchi, M. A.S. Netto , R. Buyyab, "Big Data Computing And Clouds: Trends and Future Directions," Journal of Parallel and Distributed Computing, Vol. 75, Issue 13, pp. 156-175, 2014.
[35] V.N. Inukollu, S. Arsi, S. R. Ravuri, "Security Issues Associated With Big Data in Cloud Computing", International Journal of Network Security and Its Applications, Vol. 6, Issue 3, 2014.
[36] Shah, Annappa and K. C. Shet, "Design An Efficient Big Data Analytic Architecture For Retrieval Of Data Based On Web Server In Cloud Environment", International Journal on Cloud Computing: Services and Architecture, Vol. 4, Issue 2, 2014.
[37] S. Suthaharan, "Big Data Classification: Problems And Challenges In Network Intrusion Prediction With Machine Learning," in Big Data Analytic workshop, in conjunction with ACM Sigmetrics, Vol. 41, Issue 4, pp. 70-73, 2013.
[38] B.R. Prakash, M. Hanumanthappa, "Issues and Challenges in the Era of Big Data Mining", International Journal of Emerging Trends and Technology in Computer Science, Vol. 3, Issue 4, 2014.
[39] Agrawal , S. Das , A. E. Abbadi, "Big data and cloud computing: new wine or just new bottles?", Proceedings of the VLDB Endowment, Vol. 3, Issue 1-2, pp. 1647 -1648, 2010.