

# Challenges within Big Data and Big Data Analysis

Patil Suvarna C.<sup>1</sup> Chavan Dhanshree S.<sup>2</sup> Pawar Trupti B.<sup>3</sup> Kharat Tejuswini A.<sup>4</sup> Prof. Ashish N. Patil<sup>5</sup>

<sup>1,2,3,4</sup>Student <sup>5</sup>Professor

<sup>1,2,3,4,5</sup>Department of Computer Science & Engineering

<sup>1,2,3,4,5</sup>AGTI's Dr. Daulatrao Aher College of Engineering, Karad, Shivaji university Kolhapur, Maharashtra INDIA

**Abstract**— This paper gives overviews on the salient features of Big Data analysis and its challenges. We face daily challenges in data network like where to put it, data complexity, data management etc. It is difficult to process and store a large amount of data in Relational Database management system (RDBMS). Methods for querying and mining Big Data are fundamentally different from traditional statistical analysis on small samples. Big Data is often noisy, dynamic, heterogeneous, inter-related and untrustworthy. Nevertheless, even noisy Big Data could be more valuable than tiny samples because general statistics obtained from frequent patterns and correlation analysis usually overpower individual fluctuations and often disclose more reliable hidden patterns and knowledge. Big Data brings new opportunities to modern society and challenges to data scientists.

**Key words:** Apache; Data Analysis; Hadoop

## I. INTRODUCTION

Big data is used to characterize data that is high volume, high velocity, and high variety; mandatory new technologies and techniques to capture, store, and analyze it; and is used to increase decision making, provide insight and discovery, and support and improve processes.

Big Data is about vast amounts of information. Specifically, it focuses on information sets that are too large to handle in the usual manner. As usual, we mean that they can't be processed by everyday applications, like Microsoft Access or Excel. Unfortunately, even with powerful processors churning away, these applications tend to get bogged down. Add the fact that the size of the information grows each year, and you have a recipe for problems. To get an idea of what we're talking about, consider the amount of information the Internal Revenue Service (IRS) processes. It's a wonder we get our tax returns in the time frames we do.

Thus in nutshell, big data is large datasets, the category of computing strategies and technologies that are used to handle large dataset. In this context, "large dataset" means a dataset too large to reasonably process or store with traditional tooling or on a single computer. This means that the common scale of big datasets is steadily shifting and may vary incomparable from organization to organization.

## II. BACKGROUND

A database is an standardized collection of data. [1] A relational database, more hard, is a collection of strategy, tables, queries, reports, views, and other elements. Database designers typically coordinate the data to model manner of reality in a way that supports processes requiring information, such as (for example) modeling the availability

of rooms in hotels in a way that supports finding a hotel with vacancies.

- 1) A database management system (DBMS) is a computer software application that relate with end-users, other applications, and the database itself to catch and resole data. A general-purpose DBMS own the definition, creation, querying, update, and administration of databases.
- 2) Data are simply facts or figures — bits of information, but not information itself. When data are processed, interpreted, organized, structured or presented so as to make them meaningful or useful, they are called information. The information provides context for data.
- 3) For example, a list of dates — data — is meaningless without the information that makes the dates relevant (dates of a holiday).
- 4) Data and information combined together even if one is verified them as two separate words or using them, as is common today. In case they are used interchangeably on significant on the usage of "data".

### A. Examples of Data and Information

- The history of temperature inspection all over the world for the 100 years ago is data. If this data is formed and predict to find that global temperature is rising, then that is information.
- The number of visitors to a website by country is an example of data.
- Often data is required to back up a claim or conclusion (information) copied or deduced from it. For example, before a drug is validated by the FDA, the compose must conduct clinical trials and present a lot of data to determine that the drug is safe.

## III. BIG DATA ANALYTICS

Analytics are structured, or formalized, approaches to manipulating information. It covers activities like calculations, deriving new information, and documenting results, all with an eye to a particular theme. But more to the point, it does these things using a set of standardized tools. This has a couple of benefits:

The tools act as a guide for investigation. This is particularly useful in situations where you are unfamiliar with the information. Basic conclusions can be quickly drawn, which lead to more significant derivations.

The toolset is known and easy to understand. This gets you up-to-speed quickly with new information sets and allows you to progress to the next level of investigation.

The results produced by the tools act as a baseline and can be compared to external information and results. This, in turn, gives you confidence about your results and points you to more complex activities.

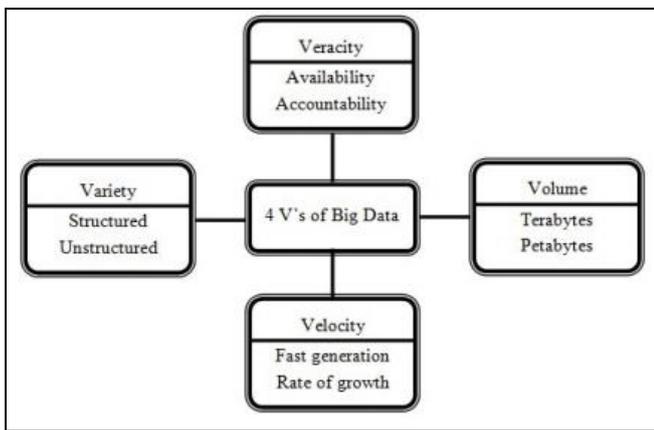


Fig. 1: characteristics of big data

#### A. Why is Big Data Important?

The importance of big data does not revolve around how much data a company has but how a company utilizes the collected data. Every company uses data in its own way; the more efficiently a company uses its data, the more potential it has to grow. The company can take data from any source and analyze it to find answers which will enable:

##### 1) Cost Savings:

Some tools of Big Data like Hadoop and Cloud-Based Analytics can bring cost advantages to business when large amounts of data are to be stored and these tools also help in identifying more efficient ways of doing business.

##### 2) Time Minimization:

The high speed of mechanisms like Hadoop and in-memory logic can easily identify new origin of data which helps businesses evaluate results immediately and make rapid decisions based on the information.

##### 3) New work Development:

By knowing the direction of customer needs and satisfaction through analytics you can create products according to the wants of customers.

##### 4) Understand the market status:

By analysis of big data you can get a better understanding of current market conditions. For example, by analyzing customers' purchasing behaviors, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors.

##### 5) Control online rank:

Big data mechanism can do sentiment study. Therefore, you can get feedback about who is saying what about your company. If you want to guide and grow the online existence of your business, then, big data tools can help in all this.

#### B. Big Data-A Competitive Advantage for Businesses

- 1) The use of Big Data is becoming common these days for the companies to outperform their peers. In most industries, existing competitors and new entrants alike will use the strategies resulting from the analyzed data to compete, innovate and capture value.
- 2) Big Data helps the institution to create new growth scope and perfectly new categories of companies that can combine and analyst industry data. These companies have sample information about the products and services, buyers and suppliers, consumer choice that can be captured and analyzed.

- 3) It also accept and expect business processes. Vendor can easily optimize their stock based on guessing models generated from the social media data, web search direction and weather forecasts.
- 4) Let our experts guide you further on the major benefits of using Big Data outlined specifically around your business model.

#### IV. TOOLS FOR BIG DATA ANALYTICS

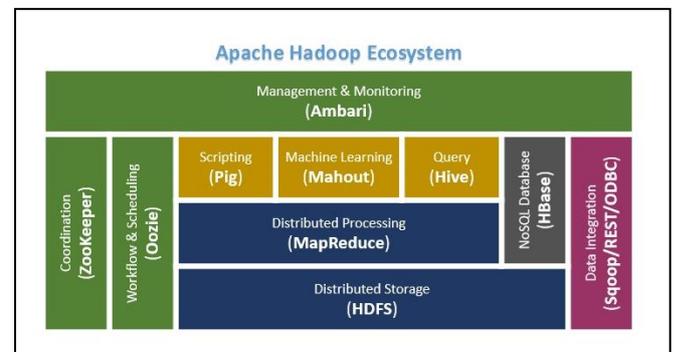


Fig. 2: Hadoop Architecture

#### A. Apache Hadoop

Apache Hadoop is a java based free software framework that can effectively store a large amount of data in a cluster. This framework runs in parallel on a cluster and has an ability to allow us to process data across all nodes. Hadoop Distributed File System (HDFS) is the storage system of Hadoop which splits big data and distribute across many nodes in a cluster. This also replicates data in a cluster thus providing high availability.

#### B. Microsoft HDInsight

It is a Big Data solution from Microsoft mechanized by Apache Hadoop which is available as a benefits in the cloud. HDInsight uses Windows Azure Blob storage as the default file system. This also provides high availability at low cost.

#### C. NoSQL

While the classic SQL can be effectively used to handle a huge amount of structured data, we need NoSQL (Not Only SQL) to handle unregulated data. NoSQL databases store unstructured data with no particular schema. Each row can have its own set of column values. NoSQL gives better work in storing a massive bulky of data. There are many open-source NoSQL DBs available to analyze Big Data.

#### D. Hive

This is a assign data management for Hadoop. This substructure SQL-like query option HiveSQL (HSQL) to in big data. This can be essentially used for Data mining purpose. This runs on top of Hadoop.

#### E. Sqoop

This is a mechanism that connects Hadoop with distinct relational databases to removal data. This can be effectively used to transfer structured data to Hadoop or Hive.

#### F. PolyBase

This attempt on primary of SQL Server 2012 Parallel Data Warehouse (PDW) and is used to entrance data stored in PDW. PDW is a data warehousing appliance constitution for

processing any volume of relational data and provides an alliance with Hadoop allowing us to access non-relational data as well.

#### G. Big data in EXCEL

As several people are comfortable in doing analysis in EXCEL, a famous tool from Microsoft, you can also connect data reserved in Hadoop using EXCEL 2013. Hortonworks, which is essentially working in providing Enterprise Apache Hadoop, provides an option to access big data stored in their Hadoop platform using EXCEL 2013. You can use Power View feature of EXCEL 2013 to easily summarise the data. (More information).

Similarly, Microsoft's HDInsight allows us to connect to Big data stored in the Azure cloud using a power query option. (More information).

#### H. Presto

Facebook has advanced and freshly open-sourced its Query engine (SQL-on-Hadoop) named Presto which is constitution to handle petabytes of data. Unlike Hive, Presto does not depend on MapReduce technique and can quickly recover data.

### V. APPLICATIONS OF BIG DATA

#### A. Big Data in Healthcare

The big data is in spread use in the patch of medicine and healthcare. As the technology accession the worth of healthcare is also increasing more and more. Big data is a tremendous serving hand in this problems. It is a great help for even physicians to keep indication of all the patients' history. The attachment to the patient's history can be approached only by the patient and his particular physician.

Once a patient gets estimate his name and his data will be gathered in the database carefully forever and whenever required, the doctor can have a view of it. A huge number of medical devices are there which are big data adapt. Today data is used to such an extent that doctor prescribes the medicines without even visiting the patient by knowing the heartbeat and temperature through the heart and temperature observing watch fitted on the patient's hand that stays in a remote place.

Nanorobots are mini robots that are actuality established which will spread the liberty in the human's body by fighting with bacteria and other harmful germs. They have their owned sensors and will be great for delivering cancer. Nanobots are great biotech robots that will be used in carrying oxygen, destroy germs, and renovate tissues.

#### B. Big Data Contributions to Public Sector

In the mutual zone, the major battle are the blend and capacity of the big data from corner to corner of several mutual zone units and allied intercourse. Big data bring a large range of efficiency to the government sectors including the power investigation, deceit recognition, fitness interconnected exploration, economic promotion investigation, and ecological fortification.

Big data is straight used to consider the food based pollution by the FDA. Big data conclusion are fast which outputs to quicker welfare. Also in the survey of a huge volume of public complaints uses the big data analytics.

This someday are apply in the system of health check demography in urgency and resourcefully for quicker judgment manufacture and to become alert of suspicious or falsified declarations.

#### C. Big Data Contributions to Learning

Big data has great significance in the teaching world too. Today approximately every procedure of information is present online. Along with the online information, there are many examples of the use of big data in the research industry. Employment named as the Bubble Score allow teachers to transmit multiple-choice estimate through mobile devices and score up paper tests over the cameras of the mobile phones. machinery like this usually assists teachers to send out the outputs to rank books and trail development all along definite characteristics.

#### D. Adaptive learning:

Another then just reorganization system work and the classify development, data-driven classrooms free up the understanding of what children learn when they study it and to what height. establishment produce intelligence courses that use big-data-fuelled guessing logic to locate what a learner is research and what ingredient of a lecture plan most excellently ensembles them at those situations.

#### E. Problem control:

Occasionally, students acknowledge his friend's homework on behalf of his own. In that position, alternatively of getting the abuse, he gets reorganization and the other innocent student gets the punishment. So in this position, big data inspire the cross-checks of the homework in order to find out whose writing matches with the homework writing.

#### F. Big Data Contributions to Insurance Services

Be unsatisfactory in modified utility, be short of becoming charging and the need of put upon services to fresh fraction and to specific market segments are some of the main challenges. Big data is the component tool that is subsistence used in the manufacturing to offer consumer insights for see-through and simpler commodities, by finding out and forecast purchaser behavior from side to side information obtained from internet websites including the social media as well as CCTV video recording.

The big data as well prepare the better customer defense from insurance office. In the demand administration, extrapolative big data occupation analytics has been applied to provide more rapid service given that enormous quantity of information can be worked on notably in the authorized period. Fraud discovery has also been repaired. In the system of huge data from abacus course and social media, real-time controlling of invitation all through the argument series is used to afford insights.

#### G. Big Data Contributions to Industrial and Natural Resources

The high insistence of the natural rise on this earth is provoke the high volume along with the velocity of big data. Equivalently, a great verity of data originate the built-up industry is unexpected. The unfamiliar data escape advanced eminence of merchandise, power competence, dependability, and improved income borderline. In the native treasure industry, big data provide for analytical

modeling to assist sense creation that is used to expend and organize large amounts of information from geographic information, graphical information, manuscript and historical statistics. Big data has as well been beat in finding the solution to the development of battle and to grow aggressive improvements in the middle of former settlements.

#### H. Big Data Contributions to Transportation

In recent times, unlimited volumes of statistics from area-oriented association networks and massive speediness statistics from telecoms have determine journey policies a lot. Horribly, inquiry of the acknowledgement passage policy has not developed yet. Consistently, transit requirement delegation is again oriented on unsatisfactory unstated fresh social media architectures.

A statistic of case of big data by the public distinct, private organization, and personal use include:

The private sector uses the big data in traffic management, direction preparation, intellectual transportation arrangements and overcrowding administration

The private sector uses the big data in income administration, industrial improvements, logistics and for reasonable benefit

Personal use of the big data comprises direction forecasting to accumulate on petroleum and period, for tour activities in seeing the sights etc.

#### I. Big Data Contributions to Banking Zones and Fraud Detection

Big data is extremely used in the cheat detection in the banking zone. In banking zone as the big data is implemented, it finds out all the mischief tasks done. It expose the abuse of credit cards, abuse of debit cards, archival of inspection tracks, enterprise credit risk treatment, business clarity, customer statistics alteration, public analytics for business, IT action analysis, and IT strategy fulfillment data. The SEC use this big data in the direction of keep a track of all the economic market movements.

They are at being using network analysis and natural speech processors to purchase illegal business activity in the economic marketplaces. Retail traders, Private and public actor banks, prevaricate funds and others in the monetary marketplace make use of big data for business analytics used in big businesses, reaction dimension, prognostic Analytics etc. In businesses, big data helps a lot in knowing the shopping patterns of customers and CRM tactics of the competitors so that they can apply them in their businesses in order to improve the sales.

### VI. CONCLUSION

In this paper, the Advanced Encryption Standard algorithm is studied and implementation is explained. AES is a symmetric block cipher that can process data blocks of 128 bits through the use of cipher keys with lengths of 128, 192, and 256 bits. An efficient FPGA implementation of the 128-bit block and 128-bit key AES algorithm are presented in this paper. The design is implemented using XC3S200 Spartan-3 FPGA device, which is based on high-performance architecture. AES is implemented in system c language. For analysis of the implementation of AES on

different platforms, we implement AES algorithm in VHDL also. With comparative analysis, we conclude that implementation of AES in VHDL is difficult as compared to system C. Also VHDL require high computing time than system C. Thus, FPGA implementation is an intermediate solution between general purpose processors (GPP) and Application specific integrated circuits (ASIC).

### REFERENCES

- [1] Mr. Ashish N. Patil , Dr. Devendrasingh M. Thakore. "A Big Data Analytics - Challenges With-in New Data, Metadata Management & Analysis Platforms" Vol. 4, Special Issue 4, January 2017
- [2] Prof. Ashish N. Patil , Anagha Ajay Jadhav. "Thrust Area in Data Science-Big Data and Data Analytics" Vol. 4, Special Issue 4, January 2017
- [3] Stein L. "The case of cloud computing in genome informatics", *Genome Biol*, 2010, vol 11, pg 207.
- [4] Donoho D. "High dimensional data analysis: the curses and blessings of dimensionality", In: *The American Mathematical Society Conference Los Angeles, CA, United States 7-12 August 2000.*
- [5] Benhadjyoussef,N., Machout,M. , "A compact 32-Bit AES design for embedded system," 7th International conference on design and technology of integrated systems in Nanoscale Era, pp.1-4, 16-18 May 2012.
- [6] "Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey". *International Journal of Network Management*. 11 March 2016. Retrieved 14 September 2016..
- [7] Trang Hoang ,Van Loi Nguyenl , "An efficient FPGA implementation of the Advanced Encryption Standard Algorithm," International Conference on Computing Communication Technologies, pp.1-6, 27-28 February, 2007.
- [8] Jaraweg,Y.,Tawalbeh,L , "AES-512:512 Bit Advanced Encryption Standard Algorithm Design and Evaluation," 7th International conference on Information assurance and security(IAS), pp.292-297, 5-8 Dec 2011.
- [9] Federal Information Processing Standards Publication 197 , "Advanced Encryption Standard (AES)," November 26, 2001.
- [10]B. Schneier, , "Applied Cryptography," 1996: Wiley, pp.784. R. Doud , "Hardware crypto solutions Boost VPN," *Electron. Eng. Times*, pp.57-64 1999.
- [11]R. Doud , "Hardware crypto solutions Boost VPN," *Electron. Eng. Times*, pp.57-64 1999.
- [12]Ayushi R.Doud , "A Symmetric Key Cryptographic Algorithm," *International Journal of Computer Applications*, Volume 1, pp No. 15, 2010.
- [13]Mr. Atul M. BorkarAyushiR. Doud , "FPGA Implementation of AES Algorithm",International Conference on Electronics Computer Technology (ICECT), pp.401-405, 2011.
- [14]Ashwini M. DeshpandeMr. Atul M. BorkarAyushiR. Doud , "FPGA Implementation of AES Encryption and Decryption ", International Conference on Control, Automation, Communication and Energy Conservation, 4th-6th June 2009.