# Big data Analytics:State of Arts Iot & cloud computing

Piyush jain[1] ,Manish joshi[2]
[1] Scholar, CCSIT, TMU, Moradabad.
[2] Assistant Professor, CCSIT, TMU, Moradabad

[1]Piyushjain517@gmail.com
[2]GOthroughmanish@gmail.com

*Abstract*— **In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets.**

***Keywords*- Big data, data mining, decision making. Big Data analytics.**

## I. Introduction

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives.

Big data is a term that refers to data sets or combinations of data sets whose size (volume), complexity (variability), and rate of growth (velocity) make them difficult to be captured, managed, processed or analyzed by conventional technologies and tools, such as relational databases and desktop statistics or visualization packages, within the time necessary to make them useful. While the size used to determine whether a particular data set is considered big data is not firmly defined and continues to change over time, most analysts and practitioners currently refer to data sets from 30-50 terabytes(10 12 or 1000 gigabytes per terabyte) to multiple petabytes (1015 or 1000 terabytes per petabyte) as big data. Figure No. 1.1 gives Layered Architecture of Big Data System. It can be decomposed into three layers, including

Infra structure Layer, Computing Layer, and Application Layer from top to bottom.

## II. Expansion of Big Data

### A. Big data and CLOUD COMPUTING

Cloud computing and big data are conjoined .Big data provide user the ability to use commodity computing to process distributed queries across multiple data sets and return resultant set .Cloud computing provides the underlying engine

through the use of Hadoop , distributed data-processing platforms. Large data sources are stored in a distributed fault tolerant data base and processed through a programming model for large data sets with a parallel distributed algorithm

In a cluster. The main purpose of data visualization is to view analytical results presented visually through different graphs for decision making. Big data utilizes distributed storage technology based on cloud computing rather than local

storage attached to a computer or electronic device .Big data evaluation is driven by fast-growing cloud-based applications developed using virtualized technologies. Therefore, cloud computing not only provides facilities for the computation and processing of big data but also serves as a service model .Map Reduce [43] is a good example of big data processing in a cloud environment; it allows for the processing of large amounts of datasets stored in parallel in the cluster. Cluster computing exhibits good performance in distributed system environments, such as computer power, storage, and network communications.

### *B. _BIG DATA AND MOBILE COMPUTING*

The rapid growth in popularity of smart phone and the development of wireless technology, mobile applications in the world are continuously expanding, and the proliferation of mobile devices and their enhanced onboard sensing capabilities are playing increasingly important role in the explosion of mobile data. Furthermore, advances in social networking and cyber-physical systems are making mobile data “big”, and consequently, bring challenges for management and processing. All these factors contribute to a possible new service paradigm: mobile big data driven service computing. Aspects of big data driven service computing include mobile big data collection and sensing, novel technologies for mobile big data transmissions (such as software-defined data transmissions and processing), and mobile big data mining (such as mobility and demographic tracing based data mining). Under the new service paradigm, mobile big data management techniques and innovative applications need to be extensively investigated in order to uncover the potential of mobile big data.

### *C. IOT AND BIG DATA*

IoT will enable big data, big data needs analytics, and analytics will improve processes for more IoT devices. IoT and big data can be used to improve various functions and operations in diverse sectors. Both have extended their capabilities to wide range of areas. The figure below shows the areas of big data produced. Some or the other way, data is produced through connected devices.

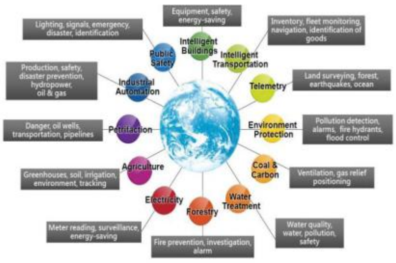
Lighting, signals, emergency, disaster, identification
Equipment, safety, energy-saving
Inventory, fleet monitoring, navigation, identification of goods
Public Safety
Intelligent Buildings
Intelligent Transportation
Production, safety, disaster prevention, hydropower, oil & gas
Industrial Automation
Telemetry
Land surveying, forest, earthquakes, ocean
Danger, oil wells, transportation, pipelines
Petrifaction
Environment Protection
Pollution detection, alarms, fire hydrants, flood control
Agriculture
Greenhouses, soil, irrigation, environment, tracking
Coal & Carbon
Ventilation, gas relief positioning
Electricity
Forestry
Water Treatment
Meter reading, surveillance, energy-saving
Fire prevention, investigation, alarm
Water quality, water, pollution, safety

Fig 1: Big data produced across various areas

The important basis behind why to implement IoT and big data are:

1. Analytical monitoring
2. More Uptime
3. Lower reject rates
4. Higher throughput
5. Enhanced safety
6. Efficient use of labor
7. Enable mass customization
8. Analyze the activities for real-time marketing
9. Improved situational alertness
10. Improved quality
11. Sensor-driven decision analytics
12. Process optimization
13. Optimized resource utilization
14. Instant control and response in complex independent systems.

The above are some possible reasons to implement IoT and Big data. As the requirements of both the technologies go hand in hand, a proper improved system is needed to overcome the challenges they pose. Many companies strive to meet the challenges and take possible steps to overcome them.

### *D. Big data AND hadoop*

Hadoop is an Apache open source framework written in Java that allows distributed processing of large dataset across cluster of computers using simple programming model Hadoop creates cluster of machines and coordinates work among them . It is designed to scale up from single servers to thousands of machines, each offering local computation and storage Hadoop consists of two component Hadoop Distributed File System(HDFS) and MapReduce Framework.

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, HDFS and numbers of various components like Apache Hive, Base and Zookeeper.

### *E. HIVE AND BIG DATA*

Scalable analysis on large data sets has been core to the functions of a number of teams at Facebook - both

engineering and non-engineering. Apart from ad hoc analysis and business intelligence applications used by analysts across the company, a number of Facebook products are also based on analytics. These products range from simple reporting applications like Insights for the Facebook Ad Network, to more advanced kind such as Facebook's Lexicon product [2].

As a result a flexible infrastructure that caters to the needs of these diverse applications and users and that also scales up in a cost effective manner with the ever increasing amounts of data being generated on Facebook, is critical. Hive and Hadoop are the technologies that we have used to address these requirements at Facebook.

### *F. SPARK AND BIG DATA*

Industries are using Hadoop extensively to analyze their data sets. The reason is that Hadoop framework is based on a simple programming model (Map Reduce) and it enables a computing solution that is scalable, flexible, fault-tolerant and cost effective. Here, the main concern is to maintain speed in processing large datasets in terms of waiting time between queries and waiting time to run the program.

Spark was introduced by Apache Software Foundation for speeding up the Hadoop computational computing software process.

Spark is not a modified version of Hadoop and is not, really, dependent on Hadoop because it has its own cluster management. Hadoop is just one of the ways to implement Spark.

### *G. PIG AND BIG DATA*

Pig Latin is a data flow language
rather than procedural ordeclarative.
• User code and existing binaries can
be included almost anywhere.
• Metadata not required, but used when
available.
• Support for nested types.
• Operates on files in HDFS.

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

## III. . STATE OF ART ANALYTICS OF BIG DATA

**Table 1**

| **Author** | **Interpretation** | **Problems Identified** | **Techniques Applied** |
|---|---|---|---|
| Hu et al [10] | 2013 | Unstructured Big Data Requires More Real-Time Analysis. | A Framework To Deteriorate Big Data Systems |
| Slavakis et al [11]. | 2013 | Managing Large Amount Of Big Data | Forth Encompassing Models |
| Srinivashan et al [12] | 2014 | Large Volumes Of Electronic Data | Two Novel Applications That Leverage Big Data To Detect Fraud, Abuse, Waste, And Errors In Health Insurance |

| | | | |
|---|---|---|---|
| | | | Claims |
| Zhang et al [13]. | 2014 | Multitasking Workloads For Big-Data. | Ordinal Optimization Using Rough Models And Fast Simulation |
| Simmhan et al [14] | 2014 | Dynamic Demand Response | Smart Grid Cyber-Physical Sagile-system<br>A Root Cause Detection Framework |
| Fu et al. [15] | 2015 | Data Structure Is Increasingly Complicated In Big Data | A Root Cause Detection Framework |
| Tan et al. [16] | 2015 | Very Large Datasets | Social Network Paradigm<br>Data Structure Is Increasingly Complicated In Big Data |
| Cevher et al. [17] | 2015 | Computational, Storage, And Communications Overheads | Approaches which are first imperative and randomized in nature. |
| Slavakis et al. [18] | 2015 | Big Data Overhead | Sentiment Analysis |
| Wu et al. [19] | 2016 | Throughput And Energy Efficiency Of Large-Scale Data Processing | Hardware-Accelerated Range Partitioner (HARP) |
| Talia et al [20]. | 2016 | Extracting Knowledge From Big Data | Smart And Scalable Analytics Services, Programming Tools, And Applications |
| Otero et al [21]. | 2016 | Power Consumption Of Big Data | Details Of Big Data Software |

## IV. CONCLUSION

Different challenges come out from the applications of Big Data Analytics in various Computing environments. but this proposed study gives more attention on many aspects which are associated with the decomposition of Big Data System in various Computing environments. The uniqueness of this paper is that this paper gives an overview of various techniques and highlights most of the significant findings of existing studies which is discussed briefly in Table I. The paper also highlights most of the significant research issues associated with the existing techniques. This survey will be beneficial for the further progress and enhancement of Big Data Analytics in various research perspectives.

## REFERENCES

[1] International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014," A review paper on Big Data and Hadoop".

[2] http://dashburst.com/infographic/big-data-volume-variety-velocity/

[3] Big Data-what is Big Data-3 Vs of Big Data- Volume, Velocity and Variety.

[4] Volume 3, Issue 10, October 2013 ISSN: 2277 128X ,International Journal of Advanced Research in ,Computer Science and Software Engineering," Big data and Methodology – A review".

[5] Sagiroglu, S.; Sinanc, D. (20-24 May 2013),"Big Data: A Review".

[6] Garlasu, D.; Sandulescu, V. ; Halcu, I. ; Neculoiu, G. ;( 17-19 Jan. 2013),"A Big Data implementation based on Grid Computing".

[7] Real Time Literature Review about the Big data.