Research Guidelines on Big Data and Data Analytics: A Survey

B.Vaishnavi
Research Student, Government Arts College, Thiruvannamalai, Tamil Nadu, India

V.Uma
Assistant Professor, Government Arts College, Thiruvannamalai, Tamil Nadu, India

C.Sunitha Ram
Assistant Professor, SCSVMV University Kanchipuram, Tamil Nadu, India

OPEN ACCESS

Volume: 6

Issue: 3

Month: Januray

Year: 2019

ISSN: 2321-788X

Received: 01.11.2018

Accepted: 03.01.2019

Published: 30.01.2019

Citation:
Vaishnavi, B., and C. Uma, V & Sunitha Ram. "Research Guidelines on Big Data and Data Analytics: A Survey." *Shanlax Internatioal Journal of Arts, Science and Humanities*, vol. 6, no. 3, 2019, pp. 7–17.

DOI:
https://doi.org/10.5281/zenodo.2550001

Abstract
Large number of devices and objects are now linked to the internet,to transmit data, and collect data back for analytics. The goal is here to utilize this data to make a positive impact on our lifestyle, energy conservation, transportation,and health. The term "Big Data" existed before IoT arrived to carry out the analytics. Themanagement of Big Data in a constantly expandingnetwork gives rise to non-trivial concerns regarding datacollection efficiency, data processing, analytics, and security. In this effort, thereforewe carry out a survey on Big Data technologies in different domains to make easy and inspire knowledge sharing across different fields. Based on the review, this work discussesanoverview, architecture, applications, challenges, techniques, methodologies, privacy and technologies, similarities and differences among Big Data technologies used in different domains, proposes how sure Big Data technology used in one realmcan be re-used in an additional area , and develops an abstract framework to outline the Big Data technologies. Then a structure may be established based on the emerging innovation behind data and analytics, managing, exploring and enabling the challenges in different task leads to the evolution of Big Data and Analytics.
Keywords: Big Data, Big Data Analytics, Internet of Things.

Introduction

The term Big Data was coined under the explosive increase of global data and was mainly used to describe the enormous datasets. These datasets are compared with traditional datasets because it does not support real- time data [i.e.,15 unstructured data].Thisleads to the evolution of Big Data [1].

In January 2007, Jim Gray, a pioneer of database software, called such transformation "The Fourth Paradigm" [2]. He also thought the only way to cope with such a paradigm was to develop a new generation of computing tools to manage, visualize, and analyze massive data. In June 2011, another milestone event occurred, when EMC/IDC published a research report titled Extracting Values from Chaos [3], which introduced the concept and potential of Big Data. This research report aroused great interest in both industry and academia on Big Data.

Big Data analytics [4] examines large amounts of data to uncover hidden patterns, correlations and other insights.Data-driven companies already using machine-generated data from the IoT to enhance customer service,generate moreyields from new products and services, optimize data and feed more data into existing analytical efforts.

Datameer is a modern BI platform, an analytics solution that helps to turn massive volumes of machine-generated sensor data into valuable, timely insights thatare powerful and yet simple for everyone to use. It runs natively on Hadoop, where we can aggregate all types of data in one place, regardless of its size.

TRUSTe[7] highlighted the fact that privacy concerns could be a significant barrier to the growth of IoT. According to the TRUSTesurvey, about 60% of internet users have simpleprivacy awareness of IoT, and theyknow that smart devices, such as smart TVs, fitness devices, and in-car navigation systems could collect personal activities data.Thefollowing points are important while designing agateway to the Big IoT data architecture. They are Rapid deployment and simple management. With the increment of new Big Data systems, the new demands on storing, analyzing,and processing different kinds of Big Data systems havespecific goals,and it must be adopted by some specific usage patterns [8].

As a result, each of them exposesdifferent interfaces,and the user must get confused to select which one is valuable for their situation. Without a flexible system for interface access and management, the applicability of the platform will be reduced much. To solve this difficult, a promising approach delivered to create a middle layer for each system to provide a unified interface [9]. However, this approach has some limitations experienced in Big Data platforms.

Minh Chau Nguyen et al. proposed a new architecture for the gateway- based access system to overcome the above challenges [10].Knowledge discovery, pattern analysis and information harvesting are the terms which are often used by the researcher inInternetof Things world. The researcher used the word “polyglot persistence” concept, in the selectionof database system in Big Data environment [12].

Several computing technologies are used 44 in wireless sensor networks like Fog, Edge and Cloud are compared with one another. The core drive of this paper is to convey the researcher to see the surprising visions of BIG IoT Technology. This paper will be helpful for all the practitioners/ researchers/academicians to choose their research path in this field. The traditional RDBMS [Relational Data Base Management System] isnot able to support heterogeneous data (i.e.,) unstructured data. All types of data including structured, semi-structured ,and unstructured data - explored from the IoT devices were handled by this new Data Base Management System – “NoSQL [Not Only SQL]” is possible in this Big Data Platforms.

Related Research Work

C.L. Philip Chen and Chun-Yang Zhang (2014), provided detailed information about the Big Data with its technologies and the different problems faced by this Big Data for handling data in various fields such as business, administration, commerce and scientific research. It also provides evidence for the opportunity to use Big Data. One leading task for the researcher is the challenges faced byevery step of the Big Data Process because which contains different techniques such as Mathematics tools, Data Analysis tools. Big Data applications are used to process information stored in the Big Data with the help of the following modes:-Batch Processing, Stream Processing and Interactive Analysis [1].

Kari Venkatram and Geetha Mary (2017), states that without analytics, there is no Big Data in this migration of technological period. Big Data characteristics are explained shortly using “V’s”, -Likewise data attributes are illuminatedusing 5C’s which are Clean,Consistent, Confirmed, Current and Comprehensive.

The NoSQL Data Base Structure stores all types of data including unstructured, semi-structured and structured data. It can be categorized based on the requirement we need, as follows: Key value store, Column-oriented database, Graph database, Document-oriented database. The term “Data Analytics” is a science used to draw insights out of the information from the data. And this analytics helps toimprove the overall decision-makingprocess in all the fields. It is divided by four types namely-Prescriptive, Predictive, Diagnosticand Descriptive. All the data arestored by Apache Hadoop architecture – A modules like a master and slave kind of architecture. Not only in Apache Hadoop,but also with Map reduces, Apache [Hive, Pig, Flume, Sqoop, Spark, Zookeeper] [4].

Ejaz Ahmed et al. (2017), the management of Big Data in a network of things arises some concern regarding data collection, efficiency, data processing, analytic, and security. All these are identified clearlywhen there is an examination made on the challenges associated with the deployment of IoT. The different group of researchers proposed the IoT based Big Data and analyticsindiversefields with some evidence. The requirements for Big Data and analytics in IoT have increased day- by- day. So there is a key – Connectivity, Storage, Quality of Service, Real-timeanalytics and Benchmark play a vital role in improving IoT services through this analytics.

The taxonomy of Big Data and analytic solutions for IoT systems are shown. It has the following attributes - Big Data sources, System components, Big Data enabling technologies, Functional elements, Analytics type. "The major sources of Big Data are IoTapplications"– this statement clearly explain the relationship between IoT and Big Data. The current IoT environment provides opportunities for current Big Data and analytics. They are Decision making, Improved Efficiency, Independence from data silos, Value added applications [13].

Mohsen Marjani et al. (2017), proposed a new architecture for Big IoT data analytics, its methods and technologies for Big Data Mining. Use cases are also presented by researchersfor easy understanding of – "How the applications are benefitted bythis new technological improvement". It discussed the existing analytical systems such as Real- time, Off-line, Memory-level, Business Intelligence ,and Massiveanalytics very efficiently. Generally, the deployment of IoT increases the amount of data in quantity and category, which in turn leads to the management of data (i.e.,) Big Data using some tools.

The Author explained about the Big Data IoT usecases (i.e.,) an application using some device. Here Sixusecases are discussed based on its benefits and whichIoT devices are used to capture data, which format the data hold [either video, audio, image, text …] and what analytics platform is used to handle the data. Opportunities provided each field to make use of this new technology for easy decision making and better processing of data. Big IoT Data arise with some challenges, even though they are solved to some extent but not fulfilled. Some issues may betackled with the help of Data Mining techniques. Data Integration leads to a lack of tools to the management of data; thisinturn leads to a complex process[11].

Sulayman K. Sowe et al. (2014) described an integrated IoT architecture with the supporting functionalities of SCN [Service Controlled Networking] in connection with Cloud Computing. The sensor data from various heterogeneous sources are managed using some protocols. The different layers present between Data-intensive research communities [i.e.,end user] SCN Model is described 91 with its architecture. The integrated IoT architecture exhibits Cloud Service models as –IaaS [Infrastructure As A Service], PaaS [Platform As A Service], SDaaS [Data / Sensor Data As A Service], SaaS [Software As A Service]. This integrated architecture for managing Data-intensive research is depicted 92 with the Japan Gigabit Network (JGN-X) layer node, middleware, physical & virtual machine servers, engines, processors and APIs. Then one case study – PM 2.5 be discussed here as an example of combining our own created sensor data with the social sensor data. This PM 2.5 sensor data is analyzed using STICKER (Spatio Temporal Information Clustering and Knowledge ExtRaction) – a Big Data Visualization tool [14].

A. Shobanadevi and G. Maragatham (2017) - reviewed the existing techniques and algorithms for IoT implementation with Hadoop and Spark technology. Hybrid data mining algorithms using the Map Reduce framework are revised. Without Mining process, the Big Data will 97 no longer possible for the organizations. Previous mining process belongs to a single main memory. But with the tremendous amount of data Parallel Computing / Mining came into existence. Clusters of computing nodes are needed when we processed some Exabyte of data and thisis possible with the deployment of some programming tools/ frameworks such as Map Reduce [15].

Perera C. Ranjan et al. describes the foremost objective of IoT is to learn more and more and to serve better to the system users. The real value of data collection comesfrom data processing and aggregation on a large scale where new knowledge can be extracted by the user. This leads to user privacy issues, which are discussed bymany researchers to produce their reportsaccordingly. TRUSTe highlighted the privacy concerns because it could be a significant barrier to the growth of IoT. Their survey revealed that 87% of internet users were concerned about the type of personal information collected. There are many challenges associated with privacy in the context of IoT which are explained here in a detailed way with some examples. They are User consent Acquisition, Control, customization and Freedom of Choice, Promise and Reality, Anonymity Technology, Security [16].

Minh Chau Nguyen et al. (2017), managing access interfaces is a chief process of analytic service development because more systems are connected to develop the process. This paper proposed a system with interface management to allow end –user to easily to do their desired functions including metadata and data access. So this system helps the platform managers, to extend and modify the access interfaces.

Multiple users belong to different fields use Big Data as a service platform. So the architecture designed herecontained eightblocks including interface management. The architecture of Gateway -based Access Interface management concept is depicted well by the researchersand described well.

Case study – ETRI Big Data Platform implemented the gateway – based access interface management system isdiscussed by the researchers. Finally,it concluded that the system could mitigate the aftermath of the heterogeneity of the interfaces provided by the several organizations [12].

Hesham El-Sayed (2017), presented the survey on Edge systems and the comparative study with cloud computing systems with IoT. Fog and Multi–cloud computing advantages are discussed. The most importantpurpose of the proposed architecture is used to provide a better Quality- of– Experience [QOE] for end users with low response time and throughput. The computing architecture for Fog Computing [FC], Cloud Computing [CC] and,MultiCloud Computing [MCC] are compared,and theirlimitations are given. Several group researchers used this Edge Computing to develop an EC based application, andthey achieved their goals. The two new emerging technologies Fog computing and Edge computing provide the better QoS [Quality of Service] to IoT applications are reachedand explained [17].

Big Data

Big Data is definedas an extreme amount of indefinite data in a variety of formats generated from a variety of sources with rapid speed in order to provide statistical results that are beneficial to the business or an industry.

Big Data Definition By Gartner - "Big Data is high-volume, velocity and variety information assets that demand cost -effective,innovativeforms of information processing for enhanced insight anddecision making"

Big Data Values

Data itself is quite often inconsequential in its own right. Big Data characteristics are defined popularly using the four V's.: volume, velocity, variety, and veracity. These four characteristics provide multiple dimensions to the value of data at hand.

Characteristics of Big Data

In 2001, Gartner accidentally aided a fall of echo with an article that forecast trends in the industry, gathering them under the headings Data Volume, Data Velocity, and Data Variety. This inflation continues in its inevitable march, and about a decade later we had 4 V's of Big Data, then 7 V's and then 10 V's. Now we are having 42 V's of Big Data and Data Science which are developed by someresearchers

Big Data and Its Evolutions

Smart Data: Beyond the volume and towards the reality

Profligatedata: Speed and agility for responsiveness

Big Data Analytics: Making smoothresolutions and predictions

Amorphousdata: Adding meaning and importance

Big Data Architecture

Big Data architecture is designed to handle the breakdown,handling, and scrutiny of data that is too large or compoundfor the outdated database systems. The edge atwhich organizations enter into the Big Data empire differs,depending on the capabilities of the users and their tools. As the trappings are advanced for working with Big Data sets, the extraction of getting values are obtainedthrough advanced analytics.

Big Data solutions involve one or more of the followingtypes of workload:

Batch processing of Big Data sources at rest. Hadoop is mainly motivated on batch data processing. An example is payroll and billing systems.

Real-time processing of Big Data in motion. Examples are Radar systems, customer services,and bank ATMs.

Interactive Exploration of Big Data: Now it becameasignificantingredient of discovery-oriented applications indiverse areas, including scientific computing, financialanalysis, evidence-based medicine and genomics Predictive analytics and machine learning.

Predictive Analytics: It is the branch of the advanced analytics which is used to make predictions about unknown future events. Example of predictive analytics is credit scoring.

Machine Learning: It is a new branch of programming and is considered an evolving technology used to enable computers to analyzea set of data and learn from the insights gathered. Examples of machine learning include classification models, recommendation engines, etc.

In future, some challenges are need to be tackled by the researcher while designing the architecture style. They are:Storing and processing of large volumes of data in a traditionaldatabase, Transforming unstructured data for inquiry and reporting.Detention, development, and evaluate unbounded streams of datain real time, or with low latency.

Comparison of Big Data Architectures

Lambda Architecture

The lambda architecture, first proposed by Nathan Marz,addresses the problem of identifying latency while large datasets are processed by some algorithmsin real time. For example, thealgorithm such as MapReduce is used to route the data in a parallel manner across this systems architecture. So this architecture creates to a new pathfor data flow. All data coming into the system goes through these two paths:

1. A batch layer (cold track) stores all the incoming data in its raw form and performs batch processing on the data.This result is stored as a batch view.
2. A speed layer (hot path) analyses data in real time. This layer is designed for low latency, at the outflow of precision.

A drawback to the lambda architecture is its complication.Processing logic appears in two different spaces — the coldand hot paths — using different frameworks. This structureleads to duplicate computation logic and the complexity ofmanaging the architecture for both ways.

Kappa Architecture

The Kappa architecture was proposed by Jay Kreps, which has the same goals as the lambda architecture, but with an important distinction: All dataflows through a single path, using thestreamprocessing system. Similar to the lambda architecture's batch layer, event data is immutable,and it is collected by, instead of a subset. The documents are ingested by a stream of events into a distributed and fault-tolerant unified log. These events are ordered by, and the current state of an event is changed only by a new experiencebeing appended.When related to speed layer, all event processing is performed on the input stream and continued as a real-time vision.

Recomputed the entire data set is required (equivalent to what the batch layer does in lambda), replay the stream, using parallelism to complete the computation in a timely fashion. Throughout the different architectures to Big Dataprocessing, the principal of data procurement boils down togathering data from distributed information sources withthe intention of storing them in accessible and capable data storage. To achieve the above goal the succeeding three main components are required:

Protocols that allow the gathering of information fordistributed data sources of any type (unstructured, semi-structured, structured data)m, Contexts with which the data is collected from the distributed sources by using different protocols, Technologies that allow the determinedstorage of the data retrieved by the Frameworks.

Phases of the Big Data Data Generation

It is the first step of Big Data. It is large scale, highly diverse and complex datasets generated through longitudinal and distributed data sources. Such data stores include sensors, videos, click streams, and all other available data sources.

Data Acquisition

It refers to the process of gathering, filtering, and cleaningdata before the data is placed into a data warehouse or any other storage solution. The acquisition of Big Data was most commonly governed by four V's. volume, velocity, variety, and value. Most data acquisition scenarios assume high-volume, high-velocity, and high-variety. But low- value data have flexibleand time-efficient gathering, filtering, and cleaning algorithms that ensure only high-value fragments of the data, which are processed by the data-warehouse.

Data Storage

Data storage is the footage (storing) of information (data)in a storage medium. Electronic data storage requireselectrical power to store and retrieve data. Data storage in a format of digital (or) machine-readable medium is sometimes called as digital data. Examples of machine-readable data on papers include Barcodes and magnetic ink character recognition (MICR).

Data Analysis

It is a process of examining, cleaning, and exhibiting data with the goal of discovering useful information, informing conclusions and supporting decision-making. Data analysis technique such as "Mining",focuses on displaying and knowledge discovery for analytical rather than purely descriptivepurposes. But business intelligence covers data analysisthat relies heavily on accumulation, which focuses mainly onbusiness statistics.

Sources of Big Data

Big Data sources are repositories of large volumes of data. Using business intelligence applications like Logi Info, users can quickly connect to and derive value from these sources. This applicationbrings more information to users' presentations without requiring the data, while that data be held in a single repository or by cloud vendor registered data store.

Examples of Big Data sources are Amazon Redshift, HP Vertica, and MongoDB. Emerging Big Data sources areanalytic/columnar data stores, NoSQL, and Hadoop data.

Some of the foremostsources of Big Data are listed as follows:

Sensors/meters and activity records from electronic devices Social interactions

Business transactions Electronic Files Broadcastings

The other examples of Big Data sources and theircorresponding mining techniques are as follows:

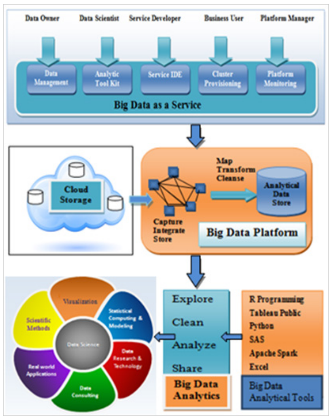
1. Social network profiles
2. Social influencers
3. Activity-generated data
4. Hadoop MapReduce application results
5. Legacy documents
6. Software as a Service (SaaS)
7. Public
8. Data warehouse appliances
9. Columnar/NoSQL data sources
10. Network and in-stream monitoring technologies

Big Data Analytics

The concept of Big Data has been around for years; mostorganizations now understand that if they capture all thedata that streams into their businesses, they can applyanalytics and get significant value from it. But even in the1950s, decades before anyone uttered the term "Big Data,"businesses were using basic analytics (essential numbers in a spreadsheet that were examined by manual process) to uncover insights and trends.

The new benefits that Big Data analytics brings to the tableare speed and efficiency. Whereas a few years ago abusiness would have gathered information, run analyticsand unearthed information that could be used by businessfor future decisions. Today that business can identify insights forimmediate decisions. The ability to work faster 211 and stay agile gives organizations a competitive edge they didn't have before. Figure 4 shows Big Data and Big DataAnalytics architecture.

Fig 4: Big Data and Big Data Analytics Framework



Why Big Data Analytics is Important?

Big Data analytics helps organizations (or) businessesto couple their dataand use it to identify new prospects, which leads to smarter business moves, more efficient operations, higher profits,andhappier customers.

Cost reduction, Faster and better decision making,new products and services, Applications of Big Data Analytics, Travel and Hospitality, Health care, Government and Retail.

Big Data Analytics and IOT

In today's world,many people would believe that theconcept of Big Data analytics was no more different fromthe notion of the Internet of things, as. "Like the fish different from the water it swims in". Both theseconcepts being big names in the industry have somesignificant differences among them, these differences onlyexist very much for the boosting of various systems andoperations of a company.

Big Data analytics solelyfall under the umbrella term for the Internet of Things, whichrefers to all those technological devices that have make the world, a closer and a better place today. One of thekey differences between these two is"fact" that these two concepts revolve around things that are somewhatdifferent from each other. When it comes to Internet ofThings, they revolve around solutions. It is an arrangement with all the various devices and services that help in offering assured solutions.

The time constraints faced by both concepts. 233 Themassive data projects consist of solutions, especially where the data is allowed to sit for almost a day or two. Then once it has settled properly, this data is then used to helpsort out things regarding all the analysis that is conducted 237. This 238 is the reason why usually it is the 234 characteristic 239 support systems which give out the longduration scenarios like capacity building, predictivemaintenance and 241 revenue protection. On the other hand Internet of Things usually ends up supporting of the real-time use scenarios such as real-time ad bidding, operationaloptimization, and detection of security breaches as well asfraud detection and so on.

Data analytics usually derives all of its data from varioussources, 240 which comes from the interaction between human generation and its information. These sourcesinclude social media activity, emails, pictures uploaded andmuch more. On the other hand, the internet of thingsdoesn't use these human-generated sources but unquestionably uses aggregation and compression of great amounts of machine-generated data that comes to us from a range ofsensors like Fitness Trackers, RFIDs and VR devices andso on.242

Data Lakes for Big Data Analytics

Data lakes are repositories where organizations strategically gather and store all the data they need toanalyze to reach a specific goal. The nature and format of the data may be semi-structured, structured, and unstructured data. The data lake is what organizations need forBig Data Analytics in a mixed environment of data. However, there are challenges to this model as well whereHadoop is a well-known solutions player and data lakes aswe know it, that theseare not an universal answer for all analytics needs.

Big Data Analytics - Use Cases

- Qubole Data Service (QDS)
- Sentiment Analysis
- 360-degreecustomerView
- Ad-hoc Analysis
- Real-time Analytics
- Multi-Channel Marketing
- Customer Micro-Segmentation
- Ad Fraud Detection
- Click Stream Analysis

Benefits of Big Data Analytics

Evaluating Big Data security analytics platforms, consider five factors that are essential to realizing the full benefits of Big Data analytics.

The Unified Data Management Platform

It is the foundation of a Big Data security analytics system that stores and queries originality data. For example, it istrying to implement distributed versions of some features of databases which include ACID transactions as well. It must balance both cost and scalability.

Support for Multiple Data Types

Big Data security analytics controls the scalability of BigData platforms with the inquiryabilities of securityanalytics and SIEM tools.

Scalable Data Ingestion

By maintaining a high write throughput of queuing data in amessage queue systems can accommodate scalable data ingestion.Instead, the data management systemmay keep afilethat acts as a buffer to hold data while it is written to disk.

Security Analytic Tools

Big Data platforms such as Hadoop and Spark are general-purpose tools and these are used forbuilding security analytical tools, but these two are not securityanalytical tools for themselves.

Compliance Reporting

It is nolonger a "nice to have" requirement. Ifit is required, review the reporting commands included with various Big Data security platforms to confirm the needs of your business are met.

The World's First Open-Source & Massively Parallel Data Platform -GREENPLUM

The Database is an advanced, fully featured, open sourcedata platform. It provides powerful and rapid analytics onpetabyte-scale data volumes. Distinctively geared towardBig Data analytics, Greenplum Database was powered by the world's most advanced cost-based query optimizerdelivering high analytical query performance on large volumes of data. The Greenplum Database project was released under the Apache 2 license. This database architecture provides automatic parallelization of all data and queries in a scale-out architecture. High-performanceloading uses MPP [Massive Parallel Processing] technology. Loading speeds scale with each additional node to greater than ten terabytes per hour, per rack. The query optimizer available in Greenplum Database isthe industry's first cost-based query optimizer for Big Dataworkloads. It can scale interactive and batch modeanalytics to large datasets in the petabytes withoutdegrading query performance and throughput. The data is accessed by the way of organizing tableor partition storage, implementation, and compression. Usershave the choice of row or column-oriented storage andprocessing for any table or partition. Apache Madlib, a library for scalable in-database analytics, which extend theSQL capabilities of Greenplum Database through user-defined functions. Access and query processing of all data is done through the external table syntax. Traditional on-premises and next-generation public data lakes are supported by this data platform.

Big Data Challenges

Hadoop is Hard

Apache Hadoop is one of the Big Data tools used to handle enormouscapacities of structured and unstructured data. Hadoopcommonly requires wide-ranging internal resources tomaintain, and many companies are left dedicating most oftheir resources to the technology rather than to the actualBig Data problem.

Scalability

With Big Data, it's crucial to be able to scale up and down on-demand. Many organizations fail to take into account how quickly a Big Data project can grow and evolve.

Continuously squeezing a project is to add extra resources cuts into time for data analysis. Big Data workloads also tend to be burst,makes it difficult toforecast where resources should be allocated. The extent of this Big Data challenge varies by the solution. A solution in the cloud will scale much at easeand faster than an on- premises solution.

Lack of Talent

Businesses are feeling the data talent shortage. Because there is a lack of data scientistsandteam members for the successful implementation of Big Data projects, and analysts who have sufficient amount of domain knowledge to identify the valuable insights.Many Big Data vendorsseek to overcome this Big Data challenge by providingtheir educational resources or by providing the bulk of the management.

Actionable Insights

Having extra data doesn't necessarily lead to actionableinsights. A key challenge for data science teams is toidentify a strong business independent and the properdatasources to collect and analyzeto meet that objective. The task doesn't stop there; however Once key patterns have been identified by the business peoples, businesses must be prepared to act and make necessary changes toderive business value from them.

Data Quality

Data quality is not a new concern, but the ability to storeevery piece of data in its unique form from a corporatethatcomplexes the problem. Common causes of cloudy datathat must be addressed by the professionalsinclude user input errors, duplicate data, and incorrect data linking. BigData algorithms can also be used to help data cleaning process.

Security

It concerns about vast lake of data secure. This is a majorBig Data challenge. It includes

User authentication for every team and team membersaccessing the data. Next it relies on restricting access based on a user's need.

Cost Management

It's difficult to project the cost of a Big Data project, andgiven how quickly they scale, can fast eat up resources. The challenge lies in taking into account allcosts of the project from acquiring new hardware/ topaying a cloud provider, to hiring additional personnel.Businesses pursuing on-premises projects must rememberthe cost of training, maintenance, and expansion. Big Data in the cloud projects must carefully evaluate theservice-level settlement with the provider to determine howusage will be billed by it and if there will be any surplus fees.

IOT and Big Data

This disruptive technology needs new infrastructures,including software and hardware applications as well as anOS; enterprises must handle the influx of data that beginsflowing in and examine it in real-time as it evolves by theminute. That is where big data arrives into the picture; bigdata analytics tools can handle large volumes of data generated from IoT devices that create acontinuous stream of information. But, todifferentiate between them, IoT provides data from whichbig data analytics can extract evidence to generate perceptions required of it.

The Role of IOT in Big Data

There are many examples of big data and IoT working welltogether to offer analysis and insight. One such example was represented by the shipping organizations. They have been utilizing big data analytics and sensor data to improveefficiency, save money and lower their environmentalimpact. They operate sensors on their delivery vehicles to monitor engine health, number of stops, mileage, miles per gallon, and speed.

IOT and big data are creating waves in big agriculture. Inthis area, the field connects systems monitors to themoisture levels and transmits this data to farmers over awireless connection. This data will enable farmers to findout when crops are reaching the optimum moisture levels.The applications of IoT and big data concepts in the field of HR enhance throughput and value. Someof the advantages here are improved the selection of talents and job matching with the required personality skills andtraits. According to a survey by business people, it is evident that both Big Data Analytics and IoT have a main role toplay in HR management.

However, IoT conducts data on a completely differentscale, so the analytics solution must accommodate its needsof processing and rapid ingestion followed by a fast andaccurate extraction.

There are many solutions available that provide near real-time analytics on large-sized datasets, and necessarilychange a full-rack database into a small server thatprocesses up to 100 TB. So small amount of hardware isneeded. The analytics database of next-generationleverages GPU technology, thus enabling even moredownsizing of the hardwares, i.e. 5 TB on a laptop or a big database in the car. It mostlyhelps IoT organizations associate the evolving number of data sets,which helps them familiarize to changing trends and attainreal-time responses, solving the challenge regarding size andcooperating on the performance.

Conclusion

"Big IoT- Data" – Make the world realistic and proximity digitized paradigm. This paper presented the overall discussion of Big Data and Data Analytics components. The architecture presented here would be better to grasp the underlying techniques using Big Data. The components are unique, and each one is described based on its underlying technology. Here all the technologies with its structures in the context of Big Data and Big data Analytics paradigm are detailed. These ideas will induce the new incoming researcher to do their research work within the field. So there is a tremendous increment in the technology must happen, which leads to overcomingall the challenges are experienced now in the BIG DATA and BIG DATA ANALYTICS ERA. Because of this new technology "Big Data", every organization must have the capability to the betterdecision-making process to improve their business in this competitive world. Not only for an organization, is every fieldbenefitted in future by this BigIoT era.

BIG DATA and BIG DATA ANALYTICS is used to handlelargeamount of heterogeneous sensor datagenerated by many devices in the digitized world. Big Data Analyticsis also capable of handling existing data mining techniques as well as new techniques from Apache Hadoop platform provided by the Big Data environment. The process ofdiscovering new procedures with a simplealgorithm is possible in this Big Data framework because by 2025 most of the devices are connected to the world of the internet.

Acknowledgments

Our sincere thanks to, who have contributed towards thedevelopment of this research paper.

References

Chen, C. P., & Zhang, C. Y. (2014). Data-Intensive Applications, Challenges, Techniques, And Technologies: A Survey On Big Data. Information Sciences, 275, 314-347.

Daniela Florescu and Donald Kossmann. Rethinking cost and performance of database systems ACMSigmod Record, 38(1):43–48, 2009.

Eric A Brewer. Towards Robust Distributed Systems. In PODC, Page 7, 2000.

Kari Venkatram, GeethaMary, A. Review On Big Data &Analytics – Concepts, Philosophy, Process And Applications, CYBERNETICS ANDINFORMATION TECHNOLOGIES • Volume 17, No 2,Sofia •2017

C. Perera, A. Zaslavsky, P. Christen And D. Georgakopoulos, "Context-AwareComputing For The Internet Of Things: A Survey," Communications Surveys Tutorials, IEEE, Vol. 16, No. 1, Pp. 414-454, 2013.

L. Atzori, A. Iera and G. Morabito, "The Internet of Things: A survey," Comput. Net.,Vol. 54, no. 15, pp. 2787-2805, Oct 2010.

TRUSTe , "Internet Of Things Industry Brings Data Explosion, But Growth Could Be Impacted By Consumer Privacy Concerns," TRUSTe Research, 29 05 2014.

Marcos D. Assunção et al., "Big Data Computing And Clouds: Trends And Future Directions," Journal Of Parallel And Distributed Computing, Vol.79, No. 80, Pp. 3- 15,May 2015.

Yue-Shan Chang, Min-Huang Ho, Shyan-Ming Yuan, "A Unified Interface For Integrating Information Retrieval," Computer Standards &Interfaces, 2001, Vol. 23 No. 4, Pp. 325-340.

Nguyen, M. C., &Won, H. S. (2017, February). Gateway-Based Access Interface Management In Big Data Platform. In Advanced Communication Technology (ICACT), 2017 19th International Conference On (Pp. 447- 450). IEEE.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., &Yaqoob, I. (2017). Big IoTData Analytics: Architecture, Opportunities, AndOpen Research Challenges. IEEE Access, 5,5247-5261.

Hwang, J. S., Lee, S., Lee, Y., &Park, S. (2015). ASelection Method Of Database System In Big Data Environment: A Case Study From Smart Education Service In Korea. International Journal of Advance Soft Computing Application, 7(1), 9-21.

Ejaz Ahmed et al. : The Role Of Big Data Analytics In Internet Of Things, Computer Networks 129 (2017) 459–471, 2017elsevier b.V.

Sowe, S. K., Kimata, T., Dong, M., &Zettsu, K.(2014, July). Managing Heterogeneous Sensor Data On A Big Data Platform: IoT Services For Data-Intensive Science. In Computer Software And ApplicationsConference Workshops (COMPSACW), 2014 IEEE 38th International (Pp. 295-300).

Shobanadevi. A. and Maragatham.G. (2017, December). Data Mining Techniques ForIotAnd Big Data—A Survey. In 2017 International Conference OnIntelligent Sustainable Systems (ICISS).

Perera, C., Ranjan, R., Wang, L., Khan, S., &Zomaya,A. (2015). Privacy Of Big Data In The Internet Of Things.

El-Sayed, H., Sankar, S., Prasad, M., Puthal, D., Gupta, A., Mohanty, M., &Lin, C. T. (2017). Edge OfThings: The Big Picture On The Integration Of Edge, IoTAnd The Cloud In A Distributed ComputingEnvironment.

Exploring Enabling Technologies and Industry Opportunities

http://epic.hust.edu.cn/minchen/min_paper/BigDataBook2014.pdf

Apache Camel. [Online]. Available: http://camel.apache.org/

MuleSoft. [Online]. Available: http://www.mulesoft.com/