Iqbal & Islam, *Journal of Applied Computing*, 1(1):14-24, 2016
e-ISSN 2456-5059
Available online at **www.zspublishers.com/jac/**

ZS Publishers

Journal of Applied Computing

REVIEW ARTICLE **OPEN ACCESS**

From Big Data to Big Hope: An outlook on recent trends and challenges

Naiyar Iqbal[1] and Mohammad Islam[2]

Department of Computer Science & IT, Maulana Azad National Urdu University (Central University), Hyderabad, Telangana-500032, India

Abstract

With the technology shift in this 21st century, Big data has its greatest impact in the society. As all knows the present era is known as information era and the information are generated and gathered by the data. Day by day data are generated rapidly via different sources, as social media, ad-hoc wireless networks, genomic data, clinical records, behavior data and many other sources. Biological data mining is used to process valuable hidden information that plays a vital role in the field of Bioinformatics and medical fields. Big data is to a great degree profitable to create efficiency in organizations and transformative leaps forward in scientific controls, which give us a considerable measure of chances to make awesome advances in numerous fields. Big data additionally emerges with numerous difficulties like, challenges in data collection, data storage, data analysis, data processing and data visualization. By the generation of huge volume of data, there are different problems and challenges arises like storage, cost, heterogeneity, uniformity. Big data is the same as adaptable analytics and the issues are basically at the application and system levels. Cloud computing has become a rapidly emerging novel computational paradigm that provide distributed, scalable accessing of the resources, and also it plays a significant role for the Big data processing. In this paper, cloud computing is explored for the solution of the Big data analytics.

Keywords: *Big Data, Big Memory, Data Analytics, Data Science, Data Mining, Cloud Computing.*

Citation:-Iqbal, N. & Islam, M. (2016).From Big Data to Big Hope: An outlook on recent trends and challenges. Journal of Applied Computing, 1(1):14-24.

Received: August 3, 2016; **Revised:** September 30, 2016; **Accepted:** September 30, 2016

Competing Interest: The author has declared that no competing interest exists.

***Corresponding Author(s):** *alnaiq@gmail.com[1], islamcs1@gmail.com[2]

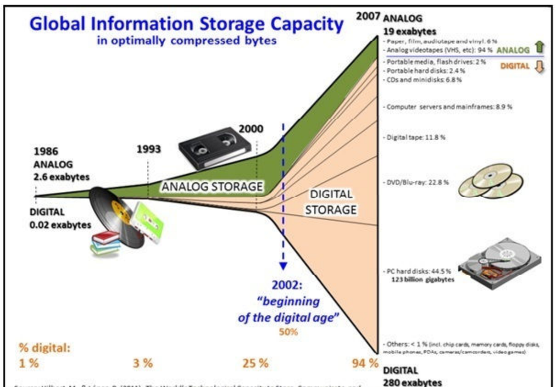
1. Introduction

Big data is an eminent research field with sound technological impact. Big data is a new and hot research topic for the current and forthcoming future. The idea of Big data is raised due to the huge amount of data produced by various sources. Big data mostly used in data warehouse, medical, education, commercial applications, social and many other fields of interest.

Heterogeneity, complexity, scale, timeliness and security issues are the hindrances in the progress of Big data. The issues begin immediately amid data gathering, when the data tsunami needed us to decide, at present in specially ad-hoc manner about what data to keep and what to dispose of and how to store what we keep dependably with the right metadata.

Big data is now become a full concept which includes different framework, tools and techniques. Big data is truly basic to the life and its developing as a standout among the most vital advancements in current world. Big data technologies are essential in giving more exact analysis, which may prompt more solid basic decision making bringing about more prominent operational effectiveness, cost effective, and decreased level of risks for the organizations. Big data implies truly a Big data, it is an accumulation of huge amount of datasets that can't be handled using customary computation and processing technique.

Big data has changed the way that it is applied in research, organizations, administrations and management. Big data alludes to gigantic measures of non-structured data delivered by superior applications falling in a large and different group of application areas, like form scientific computation, social networks, medical, education and even in e-government application.



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

fig. 1: Growth of and Digitization of Global Information Storage Capacity (Hilbert, et al., 2011).

1.1) Background of Big Data

Big data name first coined in 1990 by John Mashey in Silicon Graphics slide deck in his paper title "Big data and the Next Wave of InfraStress" (Fan, et al., 2013). NSF defines that the advantages in Big data that are necessary to analyze the huge amount of information that will be generated.

Big data basically indicated to large volume of data that can't be store and process by the use of conventional methodology within a given time period.

There are comprehensively three broad categories of Big data that is Structured data, Semi-structured Data and Unstructured data. The Structured data has a proper format related to it. Semi-structured data does not have a proper format. Whereas unstructured data does not have any format related to it.

Big Data defined by:

- ***McKinsey & company:*** *"Big data is one dataset whose size exceeds the typical database software acquisition and storage, management and analysis* (Mo, et al., 2015).*"*
- ***Victor Meyer-Schonberg's Big data:*** *"Big data means using the method of all data but not random analysis (sampling)*(Mayer-Schönberger, et al., 2013).*"*
- ***Wikipedia:*** *"Big data usually includes datasets with sizes beyond the ability of commonly used software tools to capture, curate, manage and process data within a tolerable elapsed time (Big data, Online available)."*
- ***IDC:*** *"To meet 4V (Variety, Velocity, Volume, Value) index called Big data* (Mo, et al., 2015).*"*

1.2) Bubbling "V" characteristics of Big Data:

There are many common "V" features of Big data coined by different researchers which is Volume, Variety, Velocity, Variability, Veracity and Value.

- **Volume:** The amount of data that increasingly growing day by day. As per 2012, everyday around 2.5 exabyte data are generated and that figure is twofold each forty months. More information pass over the web consistently than were put away in the whole web only twenty years prior (McAfee, et al., 2012).
- **Variety:** The different data types like textual, image, audio, video, diagram and others. The different sources like GPS signals, sensors, ad-hoc network, social networks and many more, that capture data and information updates.
- **Velocity:** The data is continuously generated at every time but only useful data are needed for the processing which gives effective information. The velocity of data production is considerably much imperative than the volume for some applications. Velocity describes as the portability of data streams. This is a challenging task to manage data generally in light of the fact that data run quick. By the use of distributed system and cloud computing, it can accomplish quick processing (Mo, et al., 2015).
- **Variability:** It is the unit of the speed of dataset. Variability explores the changing structure of data set and what is the user need to interpret that data. Irregularity of the data set can hurdle processes to manage and oversee it.
- **Veracity:** Veracity refers to the reliability and quality of the data. The data must have quality and produce trustworthy results that empower right activity with regards to end of life basic decision making.
- **Value:** Value reflects the estimation of the significance of Big data applications, which has a need worth, vulnerability and grouped qualities. A definitive target of any Big data task ought to be to produce some kind of worth for the organization doing all the investigation.

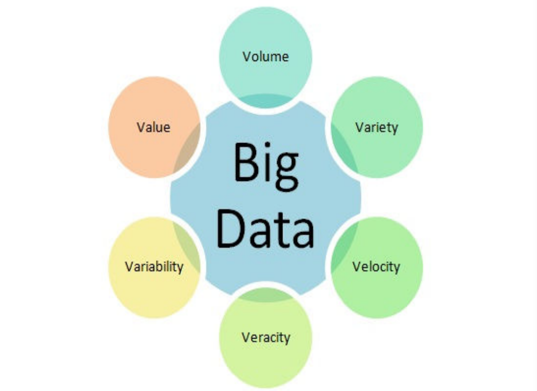


fig.2: Various V's characteristics of Big Data

1.3) Big Memory

Big memory is a combination of software and hardware approach that provides storage, accessing and management of huge data sets. The term Big memory is much similar as Big data and in a few occurrences it is a type of Big data preparing architecture implemented in memory instead of in-storage or disks. Different caches are one of the use of the big memory. BigMemory has been release recently by Terracotta, which claims that evacuate the burden that Java's garbage collector places on clients of huge Virtual Machines(Mayer-Schönberger, et al., 2013).

1.4) Data Science

Data science is an arrangement of fundamental rules which is backing and guiding the principle extraction of data and learning from data. Data mining is co-related to the data science that the real extraction of information and knowledge from data through advances that join these standards. Data science likewise is connected for general customer relationship management to investigate customer conduct keeping in mind the end goal to oversee whittling down and augment expected client value. The finance companies utilize data science for credit scoring and exchanging and in operations by means of extortion location and workforce administration.

A data science point of view furnishes experts with structure and standards, which give the data researcher a structure to efficiently treat issues of separating useful learning from data. Data science includes standards, procedures, and strategies for comprehension phenomena through the computerized analysis of data. A definitive objective of data science is enhancing basic decision making system for the enthusiasm to business.

As in section 1, it is explained the basic definition and background of Big data concept. The rest of the sections of this paper is sorted out as follows. Section 2 portrays the problem with Big data. In Section 3, we explain the new database management for Big data which is NoSQL. Further Section 4 provides some frameworks for Big data. Section 5 presents some famous used platform of Big data by different vendors. The applications area of Big data is shown in Section 6. Section 7 presents brief introduction of Big data analytics. Section 8 presents the cloud computing for Big data analytics. In section 9, Data mining and machine learning with Big data is explained. Many opportunities & challenges in Big data are reviewed in Section 10. Section 11 presents the conclusions. Table 1 shows some list of Big data application and their corresponding companies. Table 2 listed some cloud technology backup service provided by leading companies. The list of some abbreviation used in this paper is mentioned at the end.

2. Problem with Big Data

There are various problems that are faced in Big data management like Storage, processing and visualization.

Various types of data generated by different sources like social media, ad-hoc network, sensors and web application that needs to be managed, analyze, optimize and visualized.

- **Storage:** In every 3 years, the storage capacity of the information or data has roughly doubled. In numerous fields like, financial and medical data regularly be erased due to that there is no enough storage to store these data. These biological and financial data are generated in a large scale and stored at high cost, but yet overlooked ultimately.
- **Processing:** When user place a query in Big data storage, processing speed is the most interested factor on demand. Mostly it takes too much time in accessing the all the related data from all databases in less period of time.
- **Visualization:** The principle target of data visualization is to present to learning all the more naturally and successfully by utilizing distinctive diagrams. To pass on information effectively by giving information covered up in the complicated and substantial scale data sets, both appropriate structure and usefulness are essential. Present Big data representation for the most part have been insignificant exhibitions in functions, versatility and reaction time.
- **Data Management:** Big data forwards new difficulties for the management and analysis of data and further more for the entire information technology industries. Present innovations of data management frameworks are not ready to fulfill the requirements of Big data, and the expanding pace of capacity of storage limited to a great deal. So in this way a transformation re-development of data structure and storage is urgently required.
- **Security:** Most of the information resources that are made accessible and managed in cloud or distributed system have a valuable assets to their clients. So the security issue is very essential for the Big data management.

3. From SQL to NoSQL for Big Data

NoSQL is referred as Non- Relational or Non SQL database management system to store and manage the data which is something different concept to the other traditional SQL or relational database management system. NoSQL incorporates a wide assortment of various database advancements that represented in developing recent applications. At the point when contrasted with relational databases, NoSQL databases are more adaptable and give unrivaled execution.

Numerous NoSQL database innovations have superb incorporated caching ability, keeping frequent utilization of data as a part of framework storage however much as could reasonably be expected and evacuating the requirement for a different storing layer. Many NoSQL databases provide automated database replication to the management of availability in the events of run out or proper plan maintenance of the events. There are different kinds of NoSQL databases available such as graph storage, key-value store, document store and wide column store.

4. Frameworks for Big Data

There are numerous frameworks has been developed for Big data analysis that gives a new approach to conventional data analysis, like Hadoop & MapReduce, Cloud Computing, Grid Computing and others.

4.1) Hadoop & MapReduce

Hadoop is an open source system created by Doug Cutting in the year 2006 and oversaw by Apache Software establishment. Hadoop is intended to process and store efficiently a large volume of data. Hadoop structure includes with two basic parts which is HDFS and MapReduce system. The HDFS manages the storage and handles data inside the Hadoop cluster, though the MapReduce handles the computation and processing the data that is available in the HDFS.

In all the technologies, Hadoop is represented by non-relational data analysis. Hadoop has turned into the popular method for Big data processing by the characteristic of open source and simple utilizing. Hadoop moves into a standard development by the righteousness of processing for unstructured, significantly parallel and simple processing. In 2004, Google proposed MapReduce model to handle parallel processing and creating Big data, that is a linear, adaptable programming model. MapReduce is the basic of Hadoop. MapReduce is a programming model with the related computational framework that is reanimated to the primitives Map and Reduce of practical language.

MapReduce has two basic parts, the first one is map operation in which a basic function is utilized to emanate key/value sets in parallel like utilizing primary keys as a part of the relational database system. After the data to be handled is mapped into key/value assembles then the other reduce operation has apply the core processing to create results in an auspicious way (Tekiner, et al., 2013, October).

4.2) Cloud Computing

Cloud computing provides scalable service as a utility over the network. The ascent of cloud computing and cloud data storage have been an pioneer to the rise of Big data. Cloud computing has valuable features over conventional physical deployments. In spite of, cloud computing come in various structures and sometimes have to be unified with conventional models. Cloud computing utilizes perception of computing resources to run various standard virtual servers on that physical machine.

4.3) Grid Computing

Grid computing is the collaboration of distributed computer resources from many geographical area to achieve a common objective. Every resource is shared, transforming a computer network into a powerful supercomputer in the grid computing system. Individual clients can get to computers and data transparently, without knowledge of operating system, geographical area, account organization and different points of interest in grid computing environment.

The issue of Big data storage, processing and management might be resolve by the introducing distributed caching into the grid computing environment. Grid computing performs parallel processing and distributed systems which are the foundation of high performance computing (Chandhini, et al., 2013).

5. Big Data Platforms

There are some platforms mentioned for Big data management used by different vendors.

5.1) Apache Cassandra

Apache Cassandra, developed by Apache foundation, is a distributed database management system for Big data. It is an open source and free distributed database management system that provides various features to the handle of large datasets. It provides various functionalities like fault tolerance, decentralized system, consistency, query processing and scalability. Apache Cassandra gives high performance and high availability with no single point failure.

5.2) Microsoft Azure

The Microsoft Azure Big data is an open source platform for Big data based on cloud computing. It empowers to rapidly construct, deploy and oversee applications over a worldwide network system of Microsoft managed data centers. It supplies IaaS and PaaS services and also supports various frameworks and tools, programming languages. The Microsoft Azure provides many services like, data management, mobile service, computation, storage, messaging and media services.

5.3) HP Bigdata

HP's Bigdata Analytics solution contains HAVEn and Vertica. HP HAVEn platform involved HP HAVEn incorporated sorftwere, hardware and services. Structured and unstructured Big data can be analyzed to prompt effective key bits of knowledge. HP Vertica Dragline permit associations to stores the data in a savvy way, and give abilities to investigate it rapidly utilizing based on SQL instruments.

5.4) Talend Open Studio

Talend Open Studio is a flexible set of open source product to develop, test, deploy and administer the management of Big data and application combination ventures. Talend produces the main joint platform that builds data management and application integration simply by giving an integrated environment to handling with the whole life cycle crosswise over organizations.

5.5) Google BigQuery

Google BigQuery based on web service that empowers organizations to analysis on huge datasets by using framework of Google. It can be capable to analyze upto billions and more lines in every second. BigQuery is a simple and scalable to usage with the recognized SQL query. BigQuery gives developers and organizations a chance to take advantage of effective data analysis on-demand against multi terabyte datasets per seconds.

5.6) Amazon Web Service

Amazon is a web service that gives cloud based analytics services to support us to compute and analysis of any amount of data, whether our requirement is for overseen Hadoop clusters, streaming data, real-time, terabyte or more scalable warehousing of Big data or coordination.

5.7) Redhat Bigdata

Most of the Big data solutions running Linux based platform. Red Hat Enterprise Linux is a main platform for deployments of Big data. It exceeds expectations in distributed platforms and incorporates functions that address complex and major Big data requirements. Leading terrible data amount and analytic processing needs intensively an infrastructure designed for reliable, resource management, high performance, and scalable storage capability.

5.8) Cisco Bigdata

Cisco gives integrated infrastructures and analytics to bolster our large amount of data accomplice ecosystem. Cisco UCS integrated Infrastructure for Big data architecture gives a protected and adaptable infrastructure. Cisco is conveying the analytics and computation to the data to take benefit of the useful knowledge that it uncovers.

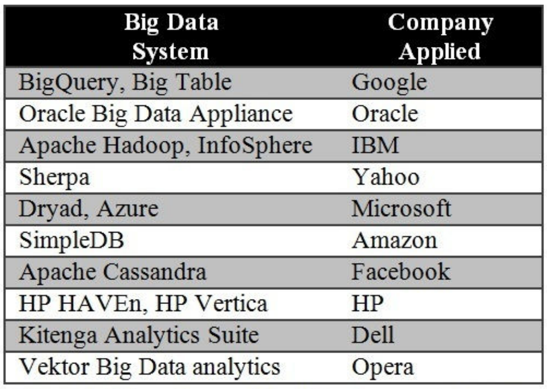
6. Application area of Big Data

At present the Big data era has discreetly slid on numerous groups, from governments and e-business to well-being associations.

6.1) Big Data for Computer Networking & IoT

Big data is used to address new concept such as cloud computing and Internet of Things (IoT). Big data addresses in all aspect of computer science for Big data management, database management, cloud computing, high performance computation, distributed system, security and privacy. Now high intelligent system is required for the management and analyzing the Big data.

Big data and the IoT co-work in the combination to each other. From a media point of view, data is the key subsidiary of gadget between availability and permits precise focusing on. With the assistance of Big data, IoT has been changes the way of media, business, organizations and governments that opening up another time of economical development and aggressiveness. The crossing point of individuals, data and expert algorithms have extensive effects on media effectiveness. The abundance of data created permits a detailed layer on the present focusing on components of the business.



Big Data System	Company Applied
BigQuery, Big Table	Google
Oracle Big Data Appliance	Oracle
Apache Hadoop, InfoSphere	IBM
Sherpa	Yahoo
Dryad, Azure	Microsoft
SimpleDB	Amazon
Apache Cassandra	Facebook
HP HAVEn, HP Vertica	HP
Kitenga Analytics Suite	Dell
Vektor Big Data analytics	Opera

Table 1: Big Data systems used by different companies
(Online Available: http://www.predictiveanalyticstoday.com/bigdata-platforms-bigdata-analytics-software/)

6.2) Big Data for Statistics

Big data is used in statistics to extract hidden patterns and the correlations with Big data for analysis. In the Big data evolution, high intelligent tools are required for investigation and perception of real time decision making to gather useful information. X-tree data structure was introduced to store high dimensional data for many Big data application (Fang, et al., 2015).

6.3) Big Data for Business Applications

The volume of business data has been increasing day by day with the emerging growth of organization. Big data analytics helps in business enterprises for identification of fraud and other risks. Big data in business analysis gives time to time and effective use of data-driven knowledge in competitive world. By the help of advanced machine learning systems to utilize the information covered up in this gigantic amount of data and it effectively enhance productivity of their evaluating procedures and publicizing effort. In the age of information, almost all big organization experiences Big data issues, particularly for multinational companies.

6.4) Big Data for Manufacturing & Production

The new emerging technologies come into existence with the moving generation, so the manufacturing company must accepts these new technological changes in the competitive world. Big data effect traditional companies for their traditional manufacturing techniques. Big data can emerge upcoming generation in manufacturing known as future manufacturing (Tekiner, et al., 2013, October)·

6.5) Big Data in Biomedical and Bioinformatics

Big data can be utilized successfully as a part of biomedical and bioinformatics. Even though complex, the big data analytics would be able to change bio-medicine, life sciences and being a healthcare services proficient or a specialist. The primary advantages of interpolating Big data analysis in customized medication incorporate reducing time while enhancing the general quality and adequacy of treating ailment. Achievement in biomedical exploration to manage the expanding measures of discards information joined with clinical data will rely on upon the capacity to translate huge data sets that are created by various developing technologies (Costa, 2014).

Day by day, there is a huge volume of biological data and information are generated related to human being and other mammals. The frequently increasing rate of huge amount of biological data or information, turns data into Big data concept. By the use of Big data with biological analytics process plays a vital role to makes medical field more significant to improve the care of the patients. Bioinformatics defined as interdisciplinary area which provides effective approach for the management, creation, storage, retrieval of biological data or information.

Today, genuine solution of the many issues in the natural field is covered up in the investigation of exponentially expanding data, purported as Big data. Big data has turned out to be presently hot and open issue for the organic group to handle, gather, store, analysis and oversee such incomprehensible measure of data and information.

4.6) Big Data for Society

Big data problem have also apply in society for public services. On the other side, the population of the world is increasing day by day, and the people at different ages need public services, like food, education, health care, public safety & security and so on. Researchers in computational science, data frameworks, sociologists, engineering, drug, and numerous different fields have been called upon to upgrade our capacity to battle brutality, terrorism, digital violations, and other digital security concerns.

7. Big data Analytics

Big data is the way toward inspecting huge data sets to reveal hidden pattern, obscure connections, market patterns, client inclinations and other helpful business data. Big data analytics is the utilization of expert analytic procedures against huge amount, different sizes and various data sets that incorporate diverse types like, unstructured or structured and batch or streaming. Experts working with Big data essentially need the learning that originates from analyzing the data.

By the use of Big data tools and software empowers an association to prepare greatly huge volumes of data that a business has assembled to make sense of which data is valuable and can be analyzed to drive more effective business choices later on. Big data analytics helps associations bridle their data and use it to recognize new way to prompts more smart business, more productive operations, more profits and more content clients. There are too many technologies that encircle Big data analytics which is used in Big data sets to extract useful information. Big data analytics is included in making that business tick by the many associations like government, health sector, travel, hospitality and others sectors.

8. Cloud computing for Big Data Analytics

In the current and forthcoming generation, cloud computing makes itself a powerful and rapid growing technology for business and information technology organizations for the solution of the Big data problems. It provides integrated environment for the parallel data processing that is helpful for the user to access resources and deployment of the programs on cloud.

Big data with the cloud computing makes a powerful platform in which Big data facilitates users to process distributed queries over various datasets and gives straightforward results. Cloud computing facilitates prime engine by the use of Hadoop and many Big data system for distributed data processing platforms (Hashem, et al., 2015). Cloud computing is an effective computational illustration for overseeing and preparing Big data storehouses.

Classification of cloud computing based on service and deployment models parameters:

a) Cloud computing based on Service model:

i. **Infrastructure as a Service (IaaS):** The IaaS supplier is in-charge of lodging, running and keeping up these administrations, by guaranteeing essential capacities like flexibility, metered service, exchange of risk and low time to advertise.

ii. **Platform as a Service (PaaS):** PaaS layer provides programming language processing infrastructure, web services, database management and operating system operation. PaaS characterizes an arrangement of instruments that give ultimate clients consistent components for making, storing, getting to and dealing with their appropriate databases on remote data servers. PaaS is the most proper computational data structure to execute huge data warehouse.

iii. **Software as a Service (SaaS):** SaaS layer performs data management, data analysis, data visualization and flexible deployment for Big data management. The applications of the SaaS layer are normally available by the user of thin client interface like a web browser (Kharche, et al., 2012).

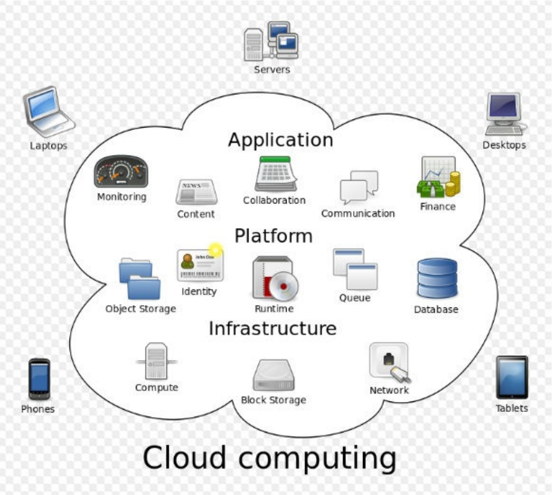


fig 3: Cloud computing metaphor: For a user, the network elements representing the provider-rendered services are invisible, as if obscured by a cloud. (Diagram showing overview of cloud computing, with typical types of applications supported by that computing model.), Created by Sam Johnston (Available online: https://en.wikipedia.org/wiki/Cloud_computing)

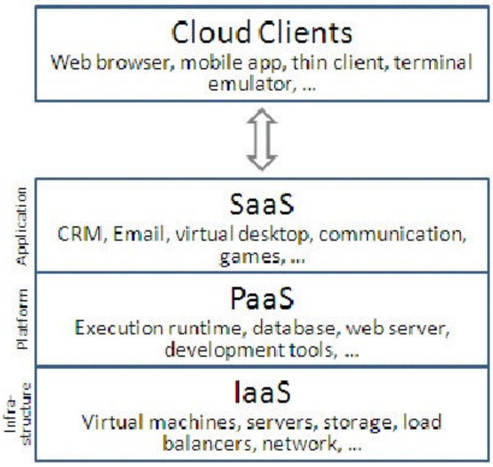

Cloud Clients
Web browser, mobile app, thin client, terminal emulator, ...
SaaS
CRM, Email, virtual desktop, communication, games, ...
PaaS
Execution runtime, database, web server, development tools, ...
IaaS
Virtual machines, servers, storage, load balancers, network, ...
Application
Platform
Infra-structure

fig 4: Cloud computing service models arranged as layers in a stack (Available online: https://en.wikipedia.org/wiki/Cloud_computing)

b) Cloud computing based on Deployment model:

i. **Public Cloud:** In public cloud, the services are open and free for the public use. The customer can be share the infrastructure and they pay on the basis of utilization of the resource or services.

ii. **Private Cloud:** Private cloud model is implemented for the specified organization. It provides more secure cloud infrastructure for dedicated client. It can be fully controlled by itself or handled by third party.

iii. **Hybrid Cloud:** It is the mixture of different kinds of clouds that combines the features of both public and private cloud infrastructures for their customer.

iv. **Community Cloud:** It is model for the similar requirements of the many individual users or organizations that share the infrastructure. This types of cloud may be handled by the community or external party.

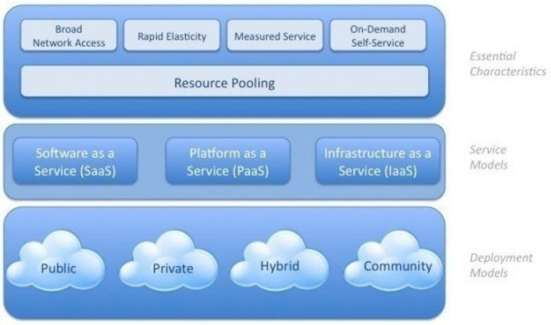


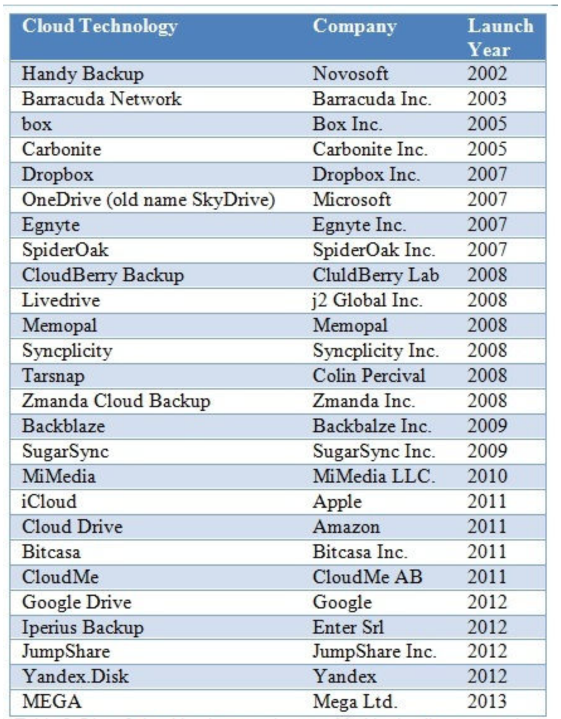
fig. 5: NIST Visual Model of Cloud Computing (Kharche, et al., 2012)

There are many leading organizations now implemented cloud computing technologies. So here the Table 2 is listed with some famous cloud technology with corresponding company and their initial release of the product.

9. Data mining and machine learning with Big Data

Data mining is the whole process of applying a computer system based approach that includes new strategies for finding learning for data and information. The capacity to extricate valuable information covered up in these data and to follow up on that learning is turning out to be progressively vital in today's competitive world. Big data mining is the capacity of separating important and useful data from the substantial datasets or surges of information, that because of its volume, speed and variability. The Big data mining upheaval is not limited to the business world because of mobile phone and handheld systems are extended in developing nations. It is evaluated that there are over five billion five billion mobile phones, and that 80% are situated in developing nations(Fan, et al., 2013). After exploratory data investigation, data scientist can be capable of figure the issue, infuse it inside the connection of an data mining, and characterize metrics for achievement.

Machine learning is the art of inspiring machines to act without being expressly customized to streamline an execution standard utilizing case data or past experience. Machine Learning systems have been produced and utilized for gathering valuable information from the data through preparing and acceptance utilizing marked datasets.



Cloud Technology	Company	Launch Year
Handy Backup	Novosoft	2002
Barracuda Network	Barracuda Inc.	2003
box	Box Inc.	2005
Carbonite	Carbonite Inc.	2005
Dropbox	Dropbox Inc.	2007
OneDrive (old name SkyDrive)	Microsoft	2007
Egnyte	Egnyte Inc.	2007
SpiderOak	SpiderOak Inc.	2007
CloudBerry Backup	CluldBerry Lab	2008
Livedrive	j2 Global Inc.	2008
Memopal	Memopal	2008
Syncplicity	Syncplicity Inc.	2008
Tarsnap	Colin Percival	2008
Zmanda Cloud Backup	Zmanda Inc.	2008
Backblaze	Backbalze Inc.	2009
SugarSync	SugarSync Inc.	2009
MiMedia	MiMedia LLC.	2010
iCloud	Apple	2011
Cloud Drive	Amazon	2011
Bitcasa	Bitcasa Inc.	2011
CloudMe	CloudMe AB	2011
Google Drive	Google	2012
Iperius Backup	Enter Srl	2012
JumpShare	JumpShare Inc.	2012
Yandex.Disk	Yandex	2012
MEGA	Mega Ltd.	2013

Table 2: List of cloud backup service provided by leading companies (https://en.wikipedia.org/wiki/Comparison_of_online_backup_services)

10. Opportunities & challenges

10.1) Opportunities in Big Data:

Those organizations who can distinguish the right framework for their Big data project and take after best practices for implementation will see a valuable competitive features. Business visionaries have additionally benefited from Big data innovation to make new items and services.

Recently, many United States government organizations like, the National Institutes of Health (NIH) and the National Science Foundation (NSF), find out that the utilities of Big data to data-profound decision making have significant impacts in their future improvements(Chen, et al., 2014).

The possibility of data generating business worth is not new; however, the compelling utilization of data is turning into the premise of competition. Big data will generally change the way organizations contend and work. Organizations that put resources into a viably separate value from their data will have a specific ideal position over their rivals.

10.2) Challenges in Big Data:

There are number of challenges in Big data that are describes in this section. Organizations would not profit from a move to utilizing bid data unless they are ready to oversee change adequately.

- **Heterogeneity:** The property "variety" of Big data is reason to the growth of heterogeneous data types from different sources, so that it leads towards the heterogeneity of the Big data. Data sources often store data of interest for the objective analytics procedures are strongly heterogeneous. Data from various sources are normally of different types and representation formats and they have inconsistent and incompatible formats.
- **Scalability:** A system is describes as scalable if it will remain effective at the point when there is a prominent increment in the quantity of resources and clients. Scalability is a capability of storage to manage rapid growth of the data in efficient way. A scalable application adapts fast growth in traffic and data volume known as scaling up, as well as, additionally adapts to decreases in demand known as scaling down.
- **Availability:** Availability is concerned with the accessibility of resources to their authorized person on demand at anytime and from anywhere.
- **Data Integrity**: Data integrity is the key concept of Big data. Its meaning that the data are only modified by authorized person and prevents from unauthorized access and misuse. In spite of, one of the essential difficulties that must be considered is to guarantee the accuracy of client data in the cloud. The user may not be access data directly, the cloud must be provide a mechanism to the client to check whether the data is managed(Hashem, et al., 2015).
- **Transformation:** This challenge refers to the exchanging of a suitable data format of the Big data, specially for the semi-structured and unstructured data.
- **Timeliness:** Timeliness is concerned with the characteristic of the Big data which is Velocity. In Big data, there is a huge volume of data, so that it takes too much time to process. But there is too many situations where result of the processing or analyzing of the data needs quickly.
- **Privacy:** Big data environment ought to be adjusted to hierarchical security and protection prerequisites. Efforts to establish safety should be actualized in order to guarantee the security of data. Protection concerns keep on hampering clients who outsource their private data into the cloud storage. This worry has gotten to be not kidding with the improvement of analysis and bid data mining, which require individual data to

deliver pertinent outcomes, for example, customized and area based service.

- **Leadership:** Companies progressively utilize Big data to reevaluate their business. Big data conveys interruption to organizations and industries that can alter the way of doing business together. The success of a business is not only depends on better and huge amount of data in this Big data evolution, but it also effected by good leadership team. The successful organizations of the new coming decade will be the ones whose pioneers can do all that while changing the way their associations settle on numerous decisions (McAfee, et al., 2012).
- **Technology:** There are numerous technologies are available in today's Big data era to handle various V's characteristics like volume, velocity, variety and others. Generally, the available tools are mostly inexpensive and open source. The Hadoop with MapReduce are most powerful framework tool available for handling Big data challenges.
- **Decision Making:** Big data presents a basic fundamental tool decision making in business. Associations are acclimated to analyzing internal data like deals, shipments, stock and inventory. The report gives understanding on their uses of Big data now and forthcoming generation. It focuses on the feature points seen and the specific issues of Big data has on-decision making for business pioneers.

11. Conclusion and future scope

Big data management community is in danger of missing the Big data train. It is unrealistic to lead Big data scrutinize viably without working together with individuals outside the data management group. The majority of Big data difficulties are being tended to by industry and lot of Big data challenges are at the development stage. Size is not only thing that matters for Big data issue, but also about to find insights from complex, noisy, heterogeneity, longitudinal and voluminous data. There are too many other challenges arise in Big data like heterogeneity, privacy, error-controlling, storage and lack of professionals, besides of huge volume of data. Data privacy is one of the major challenges of the Big data management. It issues over the whole life cycle of big data in the collection, combination, processing, analysis and uses (Mo, et al., 2015). Now the world has entered in information era with Big data. The massive changes in big data management and technology will bring about the multidisciplinary participation to hold up decision making and innovation in the services (Fang, et al., 2015).

Hadoop and MapReduce innovation used by people to get benefit of the big data applications in business, production, biomedical industry, society and different area of the interest. The cloud computing technology can be easily use for the bioinformatics big data to huge genomic sequences proficiently with appropriate speed of processing, security and economical (Nemade, et al., 2013). In this period of cloud computing it can easily store huge amount of data with the effortlessly versatility of assets in cloud Big Data prompts major challenges prompting legitimate planning to storage of the data in cloud adequately.

Taking into consideration valuable information and knowledge, it should create and make new creative systems and advancements to uncover Big data and advantage to achieve predetermined goals. So as this paper title, the Big data would be fruitful in all the sectors to manage the computation and reduce efforts and provide to useful information. So let's take a ride to the journey of Big Data towards Big Hope!

List of abbreviations

IDC	:	*International Data Corporation*
NSF	:	*National Science Foundation*
IoT	:	*Internet of Things*
NoSQL	:	*Not Only SQL/ Non-Relational*
HDFS	:	*Hadoop Distributed File System*
SaaS	:	*Software as a Service*
PaaS	:	*Platform as a Service*
IaaS	:	*Infrastructure as a Service*
UCS	:	*Unified Computing System*
LLC	:	*Limited Library Company*
NIST	:	*National Institute of Standards and Technology*

Reference

Abawajy, J. (2013). Symbioses of Big Data and Cloud Computing: Opportunities & Challenges.

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., ... & Jagadish, H. V. (2012). Challenges and Opportunities with Big Data. A community white paper developed by leading researchers across the United States. *Computing Research Association, Washington*.

Agrawal, D., Das, S., & El Abbadi, A. (2011, March). Big data and cloud computing: current state and future opportunities. In *Proceedings of the 14th International Conference on Extending Database Technology* (pp. 530-533). ACM.

Al-Hujran, O., Wadi, R., Dahbour, R., Al-Doughmi, M., & Al-Debei, M. M. Big Data: Opportunities and Challenges.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., ... & Zaharia, M. (2009). Above the clouds: A berkeley view of cloud computing.

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3-15.

Big Data [Online]. Available: Wikipedia- an internet encyclopedia- https://en.wikipedia.org/wiki/Big_data

Big Memory [Online]. Available:Wikipedia- an internet encyclopedia- https://en.wikipedia.org/wiki/Big_memory

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data* (p. 1). Washington, DC: Aspen Institute, Communications and Society Program.

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly*,*56*(1), 75-86.

Castiglione, A., Gribaudo, M., Iacono, M., & Palmieri, F. (2014). Exploiting mean field analysis to model performances of big data architectures. *Future Generation Computer Systems*, *37*, 203-211.

Cattell, R. (2011). Scalable SQL and NoSQL data stores. *Acm Sigmod Record*, *39*(4), 12-27.

Chandhini, C., & Megana, L. P. (2013). Grid computing-a next level challenge with big data. *Int J Sci Eng Res*, *4*(3).

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*,*275*, 314-347.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, *36*(4), 1165-1188.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications*, *19*(2), 171-209.

Chute, C. G., Ullman-Cullere, M., Wood, G. M., Lin, S. M., He, M., & Pathak, J. (2013). Some experiences and opportunities for big data in translational research. *Genetics in Medicine*, *15*(10), 802-809.

Cloud Computing [Online]. Available: Wikipedia- an internet encyclopedia- https://en.wikipedia.org/wiki/Cloud_computing

Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. (2009). MAD skills: new analysis practices for big data. *Proceedings of the VLDB Endowment*, *2*(2), 1481-1492.

Costa, F. F. (2014). Big data in biomedicine. *Drug discovery today*, *19*(4), 433-440.

Cuzzocrea, A., Song, I. Y., & Davis, K. C. (2011, October). Analytics over large-scale multidimensional data: the big data revolution!. In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP* (pp. 101-104). ACM.

Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113.

Diebold, F. X. (2012). A Personal Perspective on the Origin (s) and Development of'Big Data': The Phenomenon, the Term, and the Discipline, Second Version.

Diebold, F. X. (2012). On the Origin (s) and Development of the Term'Big Data'.

Dumbill, E. (2013). Making sense of big data. *Big Data*, *1*(1), 1-2.

Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM sIGKDD Explorations Newsletter*, *14*(2), 1-5.

Fang, H., Zhang, Z., Wang, C. J., Daneshmand, M., Wang, C., & Wang, H. (2015). A survey of big data research. *IEEE network*, *29*(5), 6.

Fernández, A., del Río, S., López, V., Bawakid, A., del Jesus, M. J., Benítez, J. M., & Herrera, F. (2014). Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(5), 380-409.

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, *2007*, 1-16.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., ... & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156-162.

Harford, T. (2014). Big data: A big mistake?. *Significance*, *11*(5), 14-19.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98-115.

Hassan, Q. (2011). Demystifying cloud computing. *The Journal of Defense Software Engineering*, 16-21.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, *332*(6025), 60-65.

Hoi, S. C., Wang, J., Zhao, P., & Jin, R. (2012, August). Online feature selection for mining big data. In *Proceedings of the 1st international workshop on big data, streams and heterogeneous source mining: Algorithms, systems, programming models and applications* (pp. 93-100). ACM.

http://www.predictiveanalyticstoday.com/bigdata-platforms-bigdata-analytics-software/

Ji, C., Li, Y., Qiu, W., Awada, U., & Li, K. (2012, December). Big data processing in cloud computing environments. In *2012 12th International Symposium on Pervasive Systems, Algorithms and Networks* (pp. 17-23). IEEE.

Ji, C., Li, Y., Qiu, W., Jin, Y., Xu, Y., Awada, U., ... & Qu, W. (2012). Big data processing: Big challenges and opportunities. *Journal of Interconnection Networks*, *13*(03n04), 1250009.

Johnson, J. E. (2012). Big data+ big analytics= big opportunity: big data is dominating the strategy discussion for many financial executives. As these market dynamics continue to evolve, expectations will continue to shift about what should be disclosed, when and to whom. *Financial Executive*,*28*(6), 50-54.

Kharche, M. H., & Chouhan, M. D. S. (2012). Building trust in cloud using public key infrastructure. *International Journal of Advanced Computer Science and Applications*, *3*(3).

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, *5*(12), 2032-2033.

Lee, J., Lapira, E., Bagheri, B., & Kao, H. A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment.*Manufacturing Letters*, *1*(1), 38-41.

Lin, J., & Ryaboy, D. (2013). Scaling big data mining infrastructure: the twitter experience. *ACM SIGKDD Explorations Newsletter*, *14*(2), 6-19.

Malik, P. (2013). Governing big data: principles and practices. *IBM Journal of Research and Development*, *57*(3/4), 1-1.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data. *The management revolution. Harvard Bus Rev*, *90*(10), 61-67.

Miller, H. G., & Mork, P. (2013). From data to decisions: a value chain for big data. *IT Professional*, *15*(1), 57-59.

Mo, Z., & Li, Y. (2015). Research of Big Data Based on the Views of Technology and Application. *American Journal of Industrial and Business Management*, *5*(04), 192.

Nemade, P., & Kharche, H. (2013). Big data in bioinformatics & the era of cloud computing. *IOSR-JCE*, *14*, 53-56.

O'Driscoll, A., Daugelaite, J., & Sleator, R. D. (2013). 'Big data', Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, *46*(5), 774-781.

Pandey, S., & Nepal, S. (2013). Cloud computing and scientific applications—big data, scalable analytics, and beyond. *Future Generation Computer Systems*, *7*(29), 1774-1776.

Patil, T. H., & Davenport, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*.

Petland, A. (2012). Reinventing society in the wake of big data. Edge. org.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51-59.

Ranganathan, S., Schönbach, C., Kelso, J., Rost, B., Nathan, S., & Tan, T. W. (2011). Towards big data science in the decade ahead from ten years of InCoB and the 1st ISCB-Asia Joint Conference. *BMC bioinformatics*,*12*(Suppl 13), S1.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, 1-35.

Shah, N. H., & Tenenbaum, J. D. (2012). The coming age of data-driven medicine: translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association*, *19*(e1), e2-e4.

Shim, K. (2012). MapReduce algorithms for big data analysis. *Proceedings of the VLDB Endowment*, *5*(12), 2016-2017.

Simmhan, Y., Aman, S., Kumbhare, A., Liu, R., Stevens, S., Zhou, Q., & Prasanna, V. (2013). Cloud-based software platform for big data analytics in smart grids. *Computing in Science & Engineering*, *15*(4), 38-47.

Suthaharan, S. (2014). Big data classification: Problems and challenges in network intrusion prediction with machine learning. *ACM SIGMETRICS Performance Evaluation Review*, *41*(4), 70-73.

Talia, D. (2013). Toward cloud-based big-data analytics. *IEEE Computer Science*, 98-101.

Tambe, P. (2012). Big data know-how and business value. *Working paper*.

Tan, W., Blake, M. B., Saleh, I., & Dustdar, S. (2013). Social-Network-Sourced Big Data Analytics. *IEEE Internet Computing*, *17*(5), 62-69.

Tekiner, F., & Keane, J. A. (2013, October). Big data framework. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1494-1499). IEEE.

Trelles, O., Prins, P., Snir, M., & Jansen, R. C. (2011). Big data, but are we ready?. *Nature Reviews Genetics*, *12*(3), 224-224.

Triguero, I., del Río, S., López, V., Bacardit, J., Benítez, J. M., & Herrera, F. (2015). ROSEFW-RF: the winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem. *Knowledge-Based Systems*, *87*, 69-79.

Waldrop, M. (2008). Big data: wikiomics. *Nature*, *455*(7209), 22.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering*, *26*(1), 97-107.

Zhang, L., Wu, C., Li, Z., Guo, C., Chen, M., & Lau, F. C. (2013). Moving big data to the cloud: an online cost-minimizing approach. *IEEE Journal on Selected Areas in Communications*, *31*(12), 2710-2721.

Naiyar Iqbal is currently pursuing Master of Technology (M.Tech.) in Computer Science from Department of Computer Science and IT, Maulana Azad National Urdu University (MANUU), Hyderabad, Telangana, INDIA. He has completed his Bachelor and Master degree in Computer Applications. He has 4 years teaching experience. His area of research is Machine Learning, Data Mining, Artificial Intelligence, Bioinformatics.

Mohammad Islam is currently working as an Assistant Professor in Department of Computer Science and IT, at Maulana Azad National Urdu University (MANUU), Hyderabad, Telangana, INDIA. He has 9 years teaching experience. He has completed his Bachelor and Master degree in Computer Science Engineering. He is also pursuing PhD from IIT-Delhi. His area of research is Distributed System, Artificial Intelligence, Bioinformatics. He has active membership in SCI, CSTA, IACSIT, IAENG, ISTE and UACEE.