International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Big DATA Scholarly Data in Internet of Things and Cloud Computing

DR C.P . Indhumathi Asistant prof Bharathidasan Institute of Technology Campus, Anna University, Tiruchirappalli
Dr. Dahlia Sam1 Prof Dr MGR Educational and Research Institute Maduravoyal Chennai TN
Prof Dr. Brindha Tirugnanasambandam1 Prof Dr C.Senthamarai Assistant Prof M.A.M School of Engineering Siruganur Tiruchirapalli-621105

Abstract: The technology and industries is deeply structured with new opportunities for big data using internet of things and big data. Everyday a huge number of remotely sensed data in earth observation based on this we are utilising the internet era of big data i.e. remote sensing data. Information is retrieved collected analysed using high performance computing and big data

Key words: Big data, Volume, veracity, , velocity, value, high performance computing, cloud computing

Big data is large volume of data collected from various sources in wearing degrees of complexity produced at various speed that is huge in size and growing exponentially with time. In Big Data such data is so huge and complex that traditional data management tools not able to store it or handle efficiently. While handling big data begins with aggregation of organised data which is impossible to store in the memory of a single computer. Big data processing uses a programming model to access large scale data. Hadoop is the implementation of map reduce used in big data processing.

There are a lot of benefits in big data processing such as improve customer service, utilizing outside intelligence while taking decisions, reducing maintenance cost early deduction of risk to the products/ services and does providing operational efficiency. While working with big data that a new challenges like the organisations do not have enough workers or data professionals to work with the big data tools.

The convergence of key trends is the essence of computing applications to store data in the real world into computer systems in the form of data applications where large scale of data is rapidly generated, stored in computing systems. In big data application sensors are used to generate volumes of Data Expert survey talks there are 55 billion IP sensors by 2021 storing such data is expensive every year. There is a continuous data explosion of inflow of real time exponential multiplication of data the does not seen to be slowing down there is a lot of reasons for the data explosion in the business model as shown in fig 1 and fig 2 transformations in which the innovation change to the data world governed by the organisations that have traditionally created data as a compliance requirement, with supports Limited managing report requirements. The consequence is organisations data cost consumption is minimised. Innovation in emerging trend in business, globalisations is an international scale. For manufacturing and customer service, globalisation has changed the commerce of entire globe.

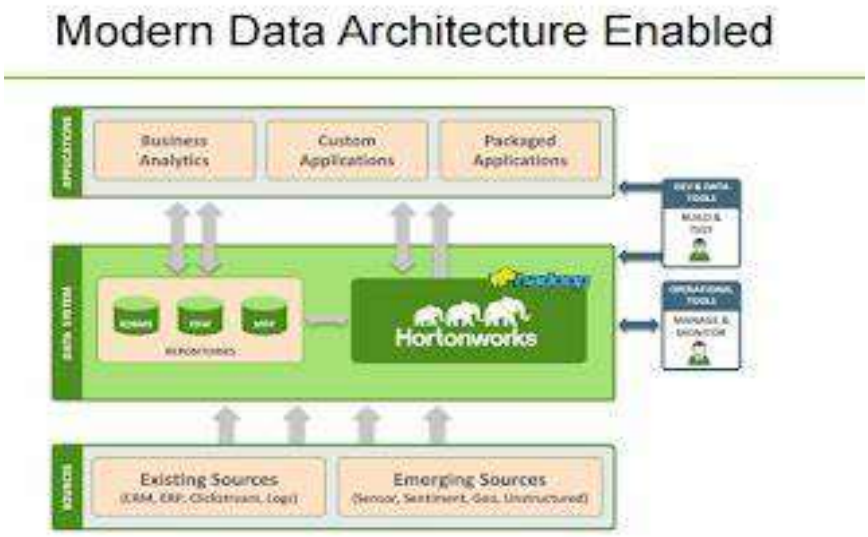


Figure 1 Big Data Architecture1

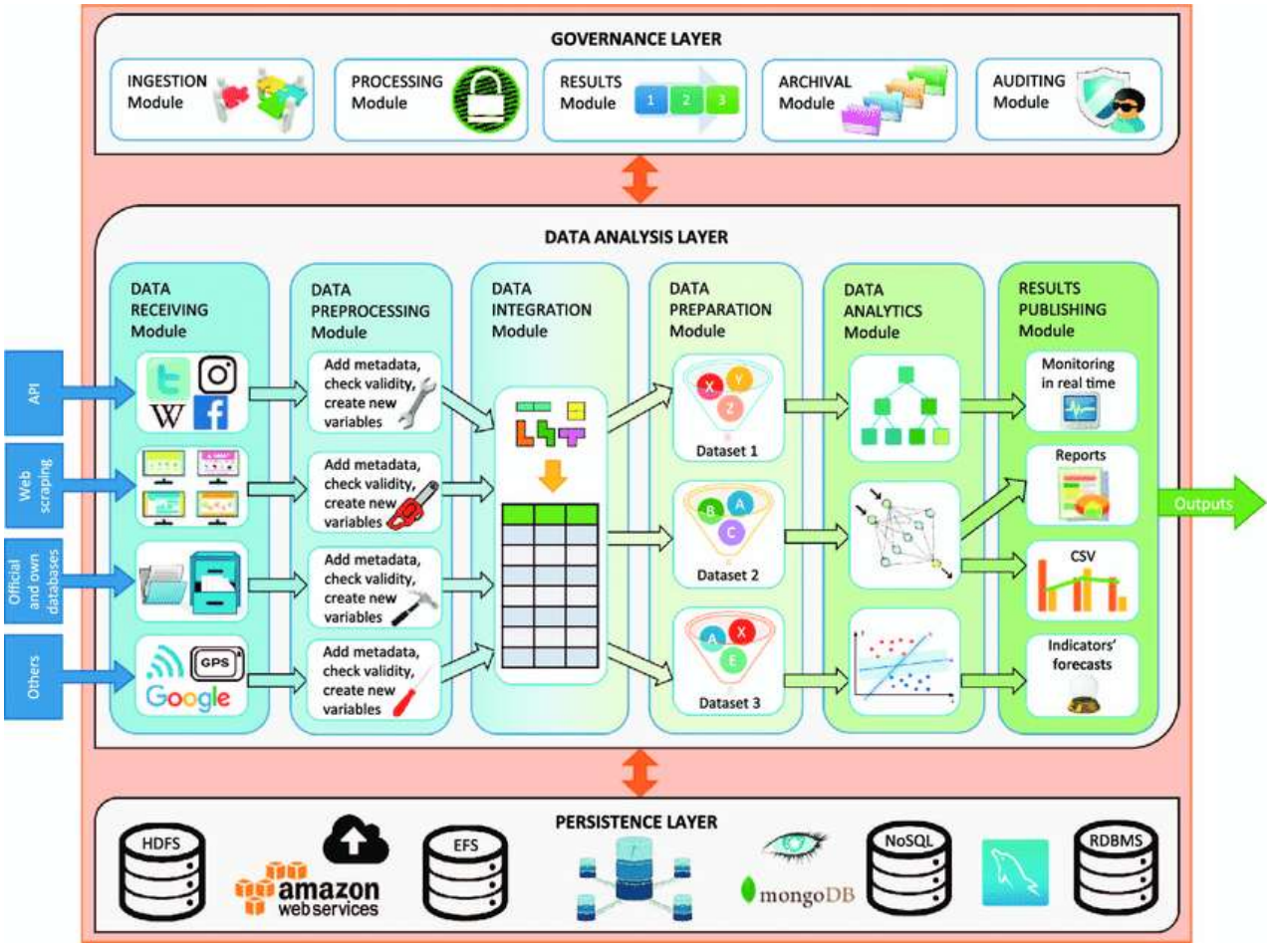


Figure 2 Big Data Architecture2

Big Data is about understanding large volumes of data collected from various sources in wearing in large quantity or degrees of data created by humans by complexity produced at as a result of introduction to new technology like say a to z gadgets from communication channels produced at various speed that is huge in size and growing voluminously or exponentially with time. In big data the data is so enormous or group of data sets and no longer a single tool or techniques rather at traditional data management tool including variety of tools

frameworks in enormous quantities, letters or symbols not able to store it or handle it efficiently.

The huge size of the big data from social network to scientific analysis in which Cloud Computing enhances business analytics by having a reduction in cost. The data exploration approach has supported provided data seizure storage and visualisation of data in large population. Cloud Computing as capabilities such as quantify ability flexibility metered pay per use sharing big data providing effective ways to analyse and help in decision making improvement in services application email Google Apps Facebook .

The cloud clients

1 software as a service

2 platform as a service

3 infrastructure as a service

The software as a service has application email Google Apps Facebook the platform services has platforms AWS elastic provided by our beanstalk Google compute engine Heroku Windows Azure. The infrastructure as a service IaAS provides infrastructure Amazon EC2 Windows Azure Google compute engine traditionally the word cloud refers to internet. Big data refers to voluminous data. It is the time sharing Technology.

Big data means handling big data beginning with a massive Collection of data that increase impossibly With market conditions business can benefit by knowing the market conditions Innovation in emerging Trends in business globalisation in an international scale in social media companies perform analysis using big data tools to get feedback about the company also improves their online presence.

To enhance customer service differentiate with the following characteristics volume, velocity, variety and value. There is no specified format like arranging row and columns that are not in use it constructs data and hence difficult to retrieve data. Example emails data images long data and videos.

with aggregation over time which is again impossible to store in the memory of a single computer thus using advanced analytic techniques programming models to access large scale data to very large heterogenous data sets ranging from terabyte to Zetabytes. Containing structured semi structure and Unstructure data from many sources.

When it accounts to big data there are a lot of benefits of big data Processing which define massive amount of data organised in sequential order and unstructured data that is encounted on daily routine. Also the insights leading to the improved mined business choices and strategic moments to improve customer service utilising outside Intelligence and taking decisions reducing maintenance cost early deduction of risk to the products services and thus providing operational efficiency.

While working with big data there is a business intelligence providing cost savings that reduce costs and improve efficiency of operations In big data the application sensors are used to generate volumes of data expert survey talks there are 55 billion IP sensors by 2021 storing such data is expensive every year. There is a discrete data explosion of inflow of data from various companies which collects data from various sources using real time in memory analytics exponential multiplication of data that does not seen to be slowing down there is a lot of reasons for business model transformations where the innovation change over time the consequence is organisation data cost is minimised.

In case of unstructured data can be arranged in that the data best suited to size is not the only problem it is complex to get the results. In various organisations more than 80% of the data are unstructured form, holds lot of information still extracting information is very big challenge. The various examples of conceptual data are machine generator data in satellite images this is the weather data scientific data this is the atmospheric data and higher energy physics, photographs and video include securities survivals and traffic video.

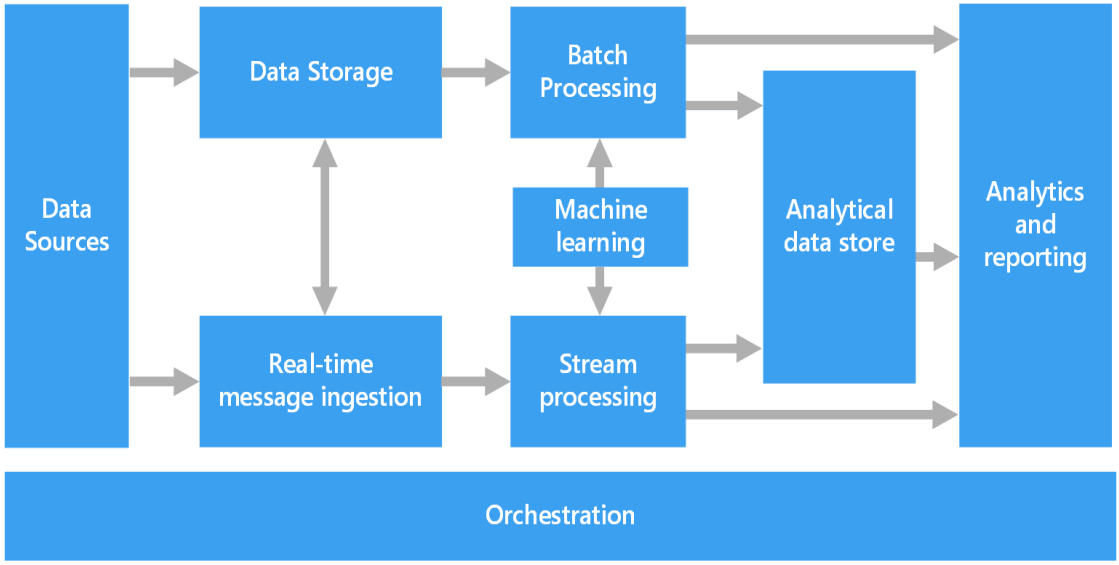


Figure 3 Big DATA architecture3

Structure data when data is arranged in rows and column format helps to application to retrieve and process data easily. There is a database management system which is used for storing the database. Structured data can be stored in the form of a fixed data call the structure data. The representation of structure data in discrete form, that is enrolled in columns format the metadeta hold the syntax and structure data document information can appear in an expected places on the document.

DATA ANALYTICS Analytics is a collection reporting and analysis of website data. There is an organisational and use a goals using website data to determine the success of failure of those goal to improve User experience. The world wide web is an open system network evolving for publishing and accessing resources services over the internet. When refining marketing campaigns, understanding website visitors, analyse website conversions, improve the website user experience, boost search engine ranking, also understand and optimise reference sources and boost online sales are the important web analytic uses.

Business purchase conversion rate use web analytics platforms. Web analytics provide insights and data that can be used to create a better User experience of website visitors. Web analytics will show us the most popular pages or website and popular parts to purchase, actually track the effectiveness of online marketing campaigns. Big data Technology defined as the technology under software utility, that is designed for analysis processing and extraction of has information from a large set of extremely Complex structures and large set of complex structures and lost data is very difficult for traditional system to deal with. Using Big Data Analytics is a way for manufacturing and customer service globalisation has changed and boost the customer acquisition and retention of the entire globe. The social media sites and networks Facebook Twitter wordpress YouTube flicker all of us ur generating data while the scientific instruments collects all sorts of data while mobile devices dealing with tracking all objects all the time and sensor technology and networks measuring all kinds of data the progress and innovation is on no

longer hinderd by the ability to collect data. When we are dealing with huge data we are not sure about the usefulness of the information collected. But by the ability to manage who is generating big data is analysed summarised visualised and the knowledge is discovered from the collected data in a timely manner and in a scaleable fashion.

To enhance customer service differentiate with the following characteristics volume velocity variety value and velocity volume refers to the data address to exabytes of existing data to process velocity refers to data in motion streaming data millions of Miliseconds to seconds to respond variety time many forms structure and structured text in multimedia veracity is the data in doubt also refers to uncertainty due to data inconsistancy and incompleteness ambiguities latency and see, deception and model approximations.

The data format supported in big data is with analysing the data with Hadoop. The format supports rich set of Meteorological elements, with variable data lengths. The default output format provided by Hadoop is text output format and it rights records is lines of the text. Output key value pairs are converted to string() method.

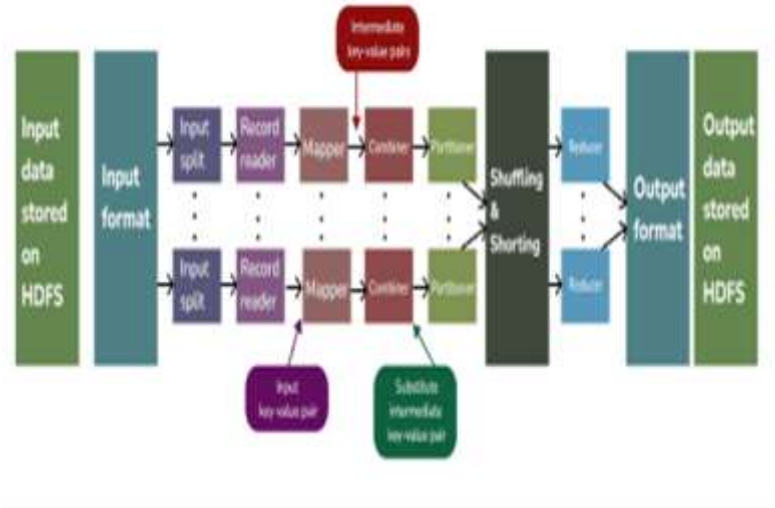


Fig4 Phases of HAdoop mapreduce - Mapreduce job execution workflow

Blocks of data are stored HDFS. These are the smallest units of data files while processing are broken into blocks and cluster is used for distribution. While analysing the data with Hadoop, Hadoop supports parallel processing which take the merit of query processing as a map reduce job. After some local small scale testing we will run in cluster of Machines. Map reduce

are broken into two phases the map phase and the reduce phase. Each face has key value pairs as input and output chosen by the programmer. Here map reduce is a method of distributing a task across multiple nodes. Each node possesses the data stored on that node to the extent possible. Map reduce is Run which consists of other phases as the following figure.

Map reduce is Run which consists of other phases as the following figure.

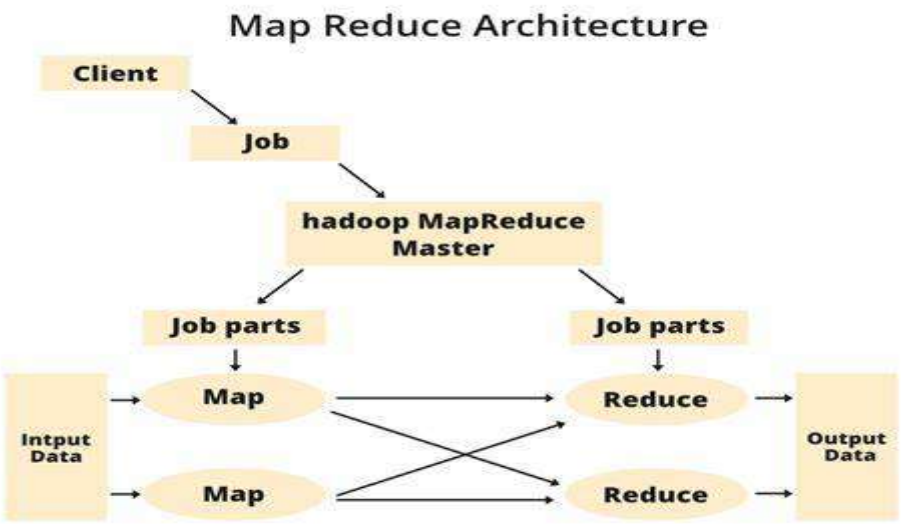


Fig 5 Phases of Hadoop map reduce

In mapping we split the input data set into chunks. Map task processes these chunks in parallel. This map outputs as input for the reduced task. Map is reduced into smaller tuple task to the final output of the framework. There are a few advantages of using map reduce which uses distributed infrastructure like CPU and storage again these are automatic parallelization and distribution of data in blocks in a distributed system and fault against failure of storage, calculating computation and network and network Infrastructures deployment monitoring and security capability and clear abstraction for programmers. Almost all the map reduce programs are written in object oriented language like Java can also be written in scripting language using the stream API of Hadoop.

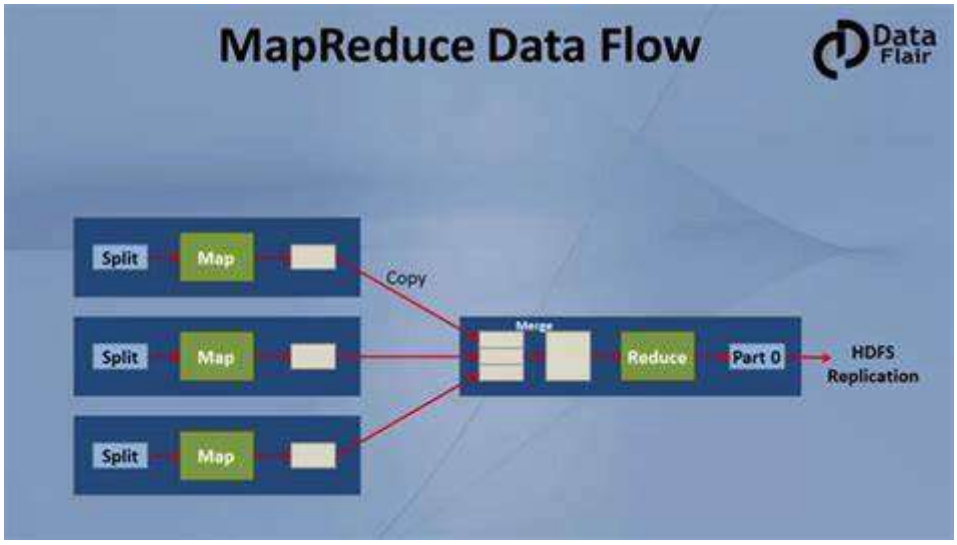
Scaling in distributed file system

For scaling out we should store the data and distributed file system HDFS allows Hadoop to move the map reduce computation to each machine. Map reduce job is the unit of work that the clients wants to be performed. It consists of the input data map produce program and configuration information. HDFS runs job by dividing it into task of which there are two types map task and reduce task. While job execution process there are two types of notes that control the process. A job tracker and number of task trackers job tracker is the one place the role of scheduling jobs and tracking jobs

Assign to the task tracker.

Job tracker is the one plays the role of scheduling jobs and tracking all jobs assigned to the task tracker. Task tracker plays the role of tracking tasks and reporting the status of task to the job tracker. Thus Hadoop divides the input to map reduce job into fixed size pieces called splits up input. Hadoop creates one map task for each split which runs the user define map function for each record in the split. This split information is used by yarn application master to try to schedule the map tasks on the same node which split data is residing thus making the task data local. Here map task write their output to the local disc not to HDFS while map output is indeterminate output and processed by reducing task to produce the final output and once a job is complete the map output is thrown away.

Fig 6 Map reduce data flow with single reduce task.



The number of reduced task is not governed by the size of the input bus is specified independently. When more number of reducer that is multiple reduces the map task partition their output, each creating one partition for each reduce task.

Map reduce data flow into multiple reduce tasks is not governed by the size of the input bus specified undependently when more number of reducer that is multiple reducers map task partition their output, each creating one partition for each reduce task. Here Hadoop allows the user to specify combiner function to be run on the map output the combiner functions output forms the input to the reduce the function. Hadoop does not provide a guarantee of how many times it will call it for a particular map output record if at all.

Hadoop streaming Adobe streaming is a utility that comes with Hadoop streaming distribution. Hadoop streaming is an application programming interface allows writing mappers and reduce in any language uses UNIX standard streams between Hadoop and the user application. Streaming is naturally suited for text processing. The data view is line oriented as key value pair the reduce function read lines from the standard input and rights output Hadoop streaming helps in real time and analysis which comes faster running on a multi-node cluster. Hadoop streaming is used by non-Java program map reduce jobs on Hadoop clusters. Hadoop streaming is flexible scalable and provides security authentication. Hadoop jobs are in such a way that it require less programming.

Hadoop pipes with Hadoop pipes we can implement applications that require higher performance in numerical calculation using C++ in map reduce

Execution of streaming and pipes

Hadoop pipes is the name of the C++ interface to Hadoop map reduce. Unlike streaming this uses Standard Input and Output to communicate with the map and reduce code. Pipes uses sockets as a channel over which the task tracker communicates with the process running the C++ map or reduce function. The figure shows the execution of streaming and pipes. The pipes utility Works by establishing a persistent socket connection on a port with the Java pipes task on one end and the external C++ process at the other. Other dedicated alternatives and implementations are also available these are the mostly built as

wrappers and JN1 based. It is however noticeable that map reduce tasks are often a smaller component to a larger aspect of changing redirecting and recurring map reduce jobs. This is usually done with the help of high level languages or APIs like pig hive and cascading which we can be used to express such data extraction and transformation problems.

Design of Hadoop distributed file system(HDFS): The Hadoop distributed file system is a distributed file system designed to run on commodity hardware. HDFS is the file system component of Hadoop. HDFS stores File system metadata and application data separately. As in the other distributed file system like GFS HDFS stores metadata on a dedicated server all the name node. Application data is stored on the other servers call data nodes. All servers are fully connected and communication with each other using TCP based protocols.

Hadoop handles large data sets running on commodity hardware scales single Apache Hadoop cluster to hundreds of nodes where a block is the minimum amount of data that it is it can read or write. HDFS blogs are 128 MB by default and this is configurable. When a file is saved in the heart HDFS the file is broken into smaller chunks or blocks.

Using Big Data Analytics is a way for manufacturing and customer service globalisation has changed and boost the customer acquisition and retention of the entire globe. The social media sites and networks Facebook Twitter wordpress YouTube flicker all of us generating data while the scientific instruments collects all sorts of data while mobile devices dealing with tracking all objects all the time and sensor technology and networks measuring all kinds of data the progress and innovation is on no longer hinderd by the ability to collect data. When we are dealing with huge data we are not sure about the usefulness of the information collected. But by the ability to manage who is generating big data is analysed summarised visualised and the knowledge is discovered from the collected data in a timely manner and in a scale able fashion.

To enhance customer service differentiate with the following characteristics volume velocity variety value and velocity volume refers to the data address to exabytes of existing data to process velocity refers to data in motion streaming data millions Mili seconds to seconds to respond variet time many forms structure and structured text in multimedia veracity is the data in doubt also refers to uncertainty due to data inconsistancy and incompleteness ambiguities latency and see, deception and model approximations.

Future research directions

Under literature they have deliberately open research concerns comprising various parameters seizure storing handling cleansing investigation gathering information secrecy of Huge data involving huge capacities of data in Big volume The research could be in

Data storage and management Data broadcast in curation Secrecy of data and security Handling data in voluminous amount and Exploration

Conclusion

Data storage with Cloud Computing is voluminous valuable and reasonable choice to moderate sized Industries with the usage of Big Data Analytics techniques dealing with background of big data and cloud computing and the challenges with the same.

References:

[1]"Discovery and Matching Numerical Attributes in Data Lakes"
Pattara Sukprasert, Pattara Sukprasert, Ryan Rossi, Fan Du, and Eunyee Koh

[2] *"Document-Level Event Argument Extraction Based on Bidirectional Span Detection"*
Yong Zhang, Feng Xiong, and Kai Zhang

[3] "ID-MixGCL: Identity Mixup for Graph Contrastive Learning"

Gehang Zhang, Bowen Yu, Jiangxia Cao, Xinghua Zhang, Jiawei Sheng, Tingwen Liu, and Chuan Zhou

[4] *"A Scalable Approach to Aligning Natural Language and Knowledge Graph Representations: Batched Information Guided Optimal Transport"*
Alexander Kalinowski, Deepayan Datta, and Yuan An

[5] "Shrinkage Denoising and Sequential State Extraction Model for Vibration Event Recognition"

Wanchang Jiang and Yuxin Jiang

[6] *"ABF-FNN: A new fuzzy neural network for predicticting coal mine gas concentration hazard"*
Yimin Sun, Xiaobo Zhang, Zhehao Zhang, Yunyang Wu, Haihao Tang, and Haonan Luo

[7] *"Multimodal Co-attention Transformer for Video-Based Personality Understanding"*
Mingwei Sun and Kunpeng Zhang

Dr C.P Indhumathi Assistant prof (Sr grade) is working as Asistant prof (Sr grade) Bharathidasan Institute of Technology Campus, Anna University, Tiruchirappalli has Research Area in Software Engineering, Software Testing, Optimization Techniques , Object Oriented Design Patterns , UI/UX Design Concepts Software Defined Network is having 15 years of experience

Authors Dr T.Brindha is currently working as an Associate professor in M.A.M School of Engineering

for Computer Science Department, Artificial Intelligence and Data Science, Trichy, Tamil Nadu, INDIA. Her interested areas include Data Base and Management Systems, Computer Networks, Object Oriented Analysis and Design, Artificial Intelligence, Big-Data, Cloud Computing, Python, Bio-informatics. She has publications in reputed journals and conferences.

Prof Dr. C.Senthamarai is currently working as an Associate professor in M.A.M School of Engineering

for Computer Science Department, Artificial Intelligence and Data Science, Trichy, Tamil Nadu, INDIA.

K.sathish kumar is working as assistant professor in the department of computer science is currently in the college