Proceedings of the 12th INDIACom; INDIACom-2018; IEEE Conference ID: 42835
2018 5th International Conference on "Computing for Sustainable Global Development", 14th - 16th March, 2018
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Various Big Data Techniques to Process and Analyze Neuroscience Data

Ashish Bhagchandani
Bachelor of Engineering Student: Information Technology Department
Gandhinagar Institute of Technology, Khatraj - Kalol Road, Moti Bhoyan, Tal. Kalol, Dist. Gandhinagar, India
ashishbhagchandani98@gmail.com

Dulari Bhatt
Assistant Professor: Information Technology Department
Gandhinagar Institute of Technology, Khatraj - Kalol Road, Moti Bhoyan, Tal. Kalol, Dist. Gandhinagar, India
dulari.bhatt@git.org.in

Madhuri Chopade
Assistant Professor: Information Technology Department
Gandhinagar Institute of Technology, Khatraj - Kalol Road, Moti Bhoyan, Tal. Kalol, Dist. Gandhinagar, India
madhuri.chopade@git.org.in

Abstract— The modern developments in neural science like neuroimaging and neuro sensing technologies have increase the size of neurological data, rate of neurological data generation, and variation in neuroscience data. These are vital role players for "Neuroscience Big data". For statistically informative datasets in terms of size, with greater time scale and colossal number of attributes, the Neuroscience community for research can develop varied type of experiments using such data [1].With the development of many data driven research techniques, the understanding for complex neurological disorder can be advance. It will also help in the field of brain networks with model long term effects of brain injury. Tools for neuroinformatics data processing and analysing are available but they are not capable to bring about huge volume of neuroscience data, which makes it hard for researchers to advance their work due to lack of capably control over this available data. So in this paper we have analysed mainly three big data techniques like map reduce, spark and pig to check their most suitability in the field of neuroscience test. Analysis shows that out of this three methods NeuroPigPen method explained in [1] is the best to deal with the challenges raised by large-scale electrophysiological signal data.

Keywords— ***Electrophysiological signal data, Neuroscience Big Data, NeuroPigPen. Neural Signal Processing, SPARK.***

I. INTRODUCTION

Neuron or nerve cell, "is an electrically excitable cell that receives, processes, and transmits information through electrical and chemical signals" [2]. It is said that, "The average human brain has about 100 billion neurons or nerve cells" [3]. To process 100 billion neurons data is obviously not a small task. Latest emerging technique like Big data can be most suitable to deal with neurons data. It is very difficult to store such data in normal hard disk as it requires space TB. So, scientists are now looking for the opportunities and challenges in big data to deal with neurons data. The metabolic reaction of ions of sodium, potassium, chloride and calcium in the neuron cell, makes a voltage gradient across its membrane [3]. The nerve impulse generates when so ever there is change in the voltage between the membrance. This nerve impulse is also known as action potential. This pulses can be captured and displayed as a waveform known as brain wave or brain rhythm[3]. It is inscribe in Brain Computer Interface And Its Types, by Anupama.H.S , N.K.Cauvery , Lingaraju.G.M, that "Brain–computer interface (BCIs) started with Hans Berger's inventing of electrical activity of the human brain and the development of electroencephalography (EEG). In 1924 Berger recorded an EEG signals from a human brain for the first time"[4]. Being a powerful communication tool between systems and user, the Brain Computer Interface (BCI) technology does not need any exterior devices or muscle involvement, for any kind of issuing commands or to complete the interaction [5].

*A. **Signal acquisition***

Signal acquisition plays vital role in neural data collection. There are many techniques as described in Table I.

TABLE I. SIGNAL ACQUISITION

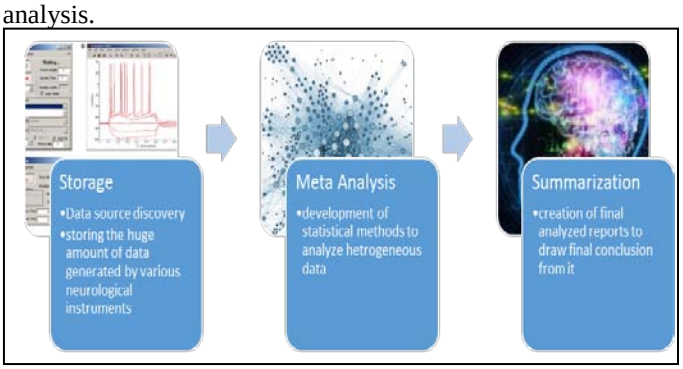
Signal acquisition	**Invasive Techniques**	**Non Invasive Techniques**
Methods	Intracortical	Magnetoencephalography (MEG)
		Functional magnetic resonance imaging (fMRI)
	Cortical Surface	Functional near-infrared spectroscopy (fNIRS)
		Electroencephalogram (EEG)

B. Role of Big Data in neuroscience

Big data is a complex form of huge is datasets. It is difficult to process it with any prevailing type of data processing application. The various type of functionalities in Big data includes, " privacy, data storage, capturing data, data analysis, searching, sharing, visualization, querying, updating, transfer, and information security" [6]. There are many big data techniques like hadoop can be used to store the data. Map

Copy Right © INDIACom-2018; ISSN 0973-7529; ISBN 978-93-80544-28-1

reduce algorithm can be useful for performing task faster. Apache spark can be used to make the system parallel and increase the speed of processing 100 times faster than map reduce. Pig can be useful to analyze the neurological data. Hive can also be useful for analyzing the data. Apache spark also uses the concept of machine learning for analysis purpose which plays important role in analysis.



II. VARIOUS TECHNIQUES OF BIG DATA

In this paper, we have identified some of the basic big data methodologies which can be used for neuroscience big data. Those all techniques are explained as below:

A. Map reduce and Hadoop for neuron data

Electroencephalogram data is an example of electrophysiological signal data, which is recorded with the help of electrodes within the skull. These electrodes play a relevant part in the analyzing of brain disorders and also it helps to diagnosis the neurological disorders. Electrophysiological data signal are also used in brain connectivity research. In neurological disorder like epilepsy, it becomes difficult for neurological scientist for clinical diagnosis as SEEG data record activities of functional brain in both temporal and spatial scales.

Analysing electrophysiological data signals like, "Electroencephalogram (EEG)" is one of the critical test or technology for diagnosing varied brain disorders [8]. It was very difficult to perceive perfect properties of Neural signals, as they are non-stationary and non-linear in identity[8], but now it has been possible after the emergence of Ensemble Empirical Mode Decomposition (EEMD) algorithm. Due to extreme complexity nature of EEMD it has become possible to process neural signal as now it is compute-intensive. It is also data intensive as,

- Use EEG signals comprise substantial datasets a zero before decimal points.
- To ensure high precision in results, a enormous number of trials have to be introduced by EEMD.

Introduction to parallel neural signal processing

Due to repetition in trials to reduce noise, EEMD commits problem of compute-intensive and data-intensive, while it analyze neural signal. Due to complex algorithm of EEMD the density and spatial scale of neural signals increases exponentially. With these two major problems parallel neural signal processing came into existence. With the help of cluster computing and grid computing in database management system the management of massive neuroimaging data takes place. This data is in form of distributed environment. DBMS plays an important role as it facilitates storing and sharing of fMRI datasets. These datasets proposes a high quality scheme for the EEG, due to ease in analysis with the help of DBMS. By this DBMS plays an important role in the part of parallel neural signal processing.

Even though it has been a consequential headway done in parallel neural signal handling system, it is tough for massive neural signals to process, a reliable and high-throughput. Failure encounters of computing elements during neural signal processing are the major drawback. The inadequate handling ability and storing size on a single server is also confined [8]. It was further mentioned by Lizhe Wang, Dan Chen, Rajiv Ranjan, Samee U. Khan, Joanna Kołodziej, and Jun Wang, in Parallel Processing of Massive EEG Data with MapReduce that, "due to this the fine grain parallel processing paradigms, such as GPGPU and OpenMP , does not provide their data processing aptness up to satisfactory. The need of new computing paradigms and platform demands parallel neural signal processing, as the ability of high throughput lacks in the distributed data intensive workflow. MapReduce is advanced data parallel processing model and Hadoop is its open-source implementation on clusters [8]. MapReduce and Hadoop have two advantages:

- Fault-tolerant storage and data processing by duplicating computing tasks on different compute nodes [8].
- High-throughput data processing via a batch processing model and a highly efficient massive file system – HDFS [8]."

B. Apache spark for neuron data

Technologies regarding scalable analysis for sizeable datasets have transpire in the field of internet computing, but are however scarcely operated in neuroimaging regardless of the existent of data and research issues in require of systematic computation tools mainly in fMRI. In fMRI datasets big data techniques are used in graph analysis, to elucidate how handling pipelines applying it can be enhanced.

Map Reduce Vs Apache Spark for fMRI Datasets

One of the most effective developments in the area of analyzing large datasets is Google's MapReduce paradigm, which is intend on datasets for logical distributed computation. These datasets are stowed in a distributed file system in a cluster environment as they are too immense to fit on a single machine.

The distributed algorithm used in large documents, "for counting the number of occurrences of words" can explain the principles of the MapReduce paradigm in perfect suitable

illustration that is, "as the name suggests, these computations are of two parts, named as Map and Reduce. Map is being operated on each node distinctly, and the Reduce part is computed at the central node by combining the distinct Map results. In the word count program of map reduce, during the map stage, associating value of a set of keys and value are generated. With the help of words being the keys and number of existence of each word in the every portion of the document is calculated on distributed file system. Reduce stage gains the general number of existences of the word in the aggregate of the dataset." There are many neurological computations which cannot be conducted using this approach. Main challenge in this approach is if system uses machine learning algorithms then it needs to access data frequently which would be very tough if it is applied using MapReduce.

By characterizing this matter and having a further extensive structure for distributed computations at great and complex datasets was the key scheme with the introduction of the Spark framework. Resilient Distributed Datasets (RDD), "was the name given to the data stored in the Hadoop distributed file system in the Spark framework" and it is not same as the files in HDFS, to maintain a high-level interface for the developer or programmer it can held the entire data in memory if space allows. Due to abstract property of distributed storage and computation on distributed dataset, the writing of distributed code is much easier.
Machine learning and graph analysis for enlacing power in specific domains, thus this is one of the key point of the Spark. The link between Spark and Scala is the toughest as the interactive shell being a Scala shell and Spark can be castoff reciprocatively from a Scala shell or also from its Application Programming Interface (API).

When it comes Big Data Technologies, parallelization of computations on enormous datasets is not straight concerned with it. Graphics Processing Units (GPUs) plays here a key role, which can help in extension of computations on a solo machine and thus making a huge modification in terms of productivity and time of computation. In the area of neuroimaging and neuroscience, "in general and functional MRI in particular, both big data frameworks and GPU acceleration can prove useful hands where increasing spatial and temporal resolutions as well as larger sample sizes lead to a rapid increase in the amount of data that needs to be processed in a typical study."

C. NeuroPigPen for neuron data

To have new apprehension into brisk of brain networks, the output from research techniques processed from advance big data can rearrange our considerate of composite neurological disorders. To discourse the encounters from enormous scale electrophysiological signal, the combination of Apache Hadoop and Pig developed a new toolkit knows as NeuroPigPen.

The information regarding datasets of fiber tract density and fiber tract orientation can be obtained with the techniques of fiber tractography by mapping brain operational networks at numerous levels of granularity and also with the high resolution magnetic resonance imaging (MRI). For example, "the patients suffering from epilepsy which is a complex neurological disorder, in which the implantation of intracranial depth electrodes at precise brain locations with the help of stereotactic placement approaches to record EEG (called SEEG) which is being increasingly used in routine clinical care for evaluation and diagnosis of patients. The customary use of signal data in both patient care and biomedical research has led to rapid buildup of large volumes of these datasets."

The method of data partitioning is very much essential for complete, "storing, processing, and analyzing" very huge volumes of composite electrophysiological and for scalable data processing applications which can run fast, the parallel computing, and distributed storage techniques that can effectually handle numerous computing nodes. Due to increase size in data, parallel processing techniques do not scale up, as they were not developed for current signal data processing tools. For great performance distributed file systems, the European Data Format, which is used to store signal data, is not fit for it. Therefore, for partitioning data over several computing nodes new signal data format is required. The time required for programming for these type of new tools compared to outdated sequential data handling software tools is also very high, for adapting new encounters in distributed and parallel computing environment.

The challenges outfaced by data and to compute intensive tasks in terms of scalability is counter balanced with the help of Stack which is an Hadoop technology. The thousands of computing nodes can parallelized the two repetitive steps of Map and Reduce which are further divided from MapReduce programming approach. The Hadoop cluster can be scaled with increasing amount of data with the help of Hadoop implementation, which is also similar to Google's MapReduce. For scientific computing, many scalable software infrastructures are developed using Hadoop platform, which is a part of Apache Foundation project.

Due to multi-step data flow and advance level of complexity of parallel operations in scalable data management tools, the use of Hadoop MapReduce is restricted to only trained developers.

The Apache Pig dataflow system was developed, to constitute numerous data handling into multi-step dataflow applications. The Apache Pig dataflow system automatically with the help of Pig complier, compiled into MapReduce task. The SQL like query constrains where also supported in Pig. The Pig Latin programs where used in describing the data processing steps in Pig. The User Defined Functions can be easily developed to handle domain specific data processing requirements. These task (UDF) or data manipulation can be easily done through Pig functions for original signal data formats or data partitioning methods.

To practice neurological signal data by previous data partitioning, data alteration and data handling performance, the NeuroPigPen toolkit was developed with several UDF's in it. In Hadoop cluster the MapReduce jobs are converted from the NeuroPigPen UDFs with the help of Pig compiler.

D. Hive for neuron

For all kinds of jobs, to practice MapReduce jobs in Hadoop requires particular JAVA functions. To handle vast amount of data, hive was settled to deliver a standard interface for database programmers that strictly match the SQL. For data stored in HDFS and data in warehousing in Hadoop, the data is fetched by Hive. Queries like commands to search, combine, or recognize the data and execute them as MapReduce Jobs within Hadoop, is guided by Hive. By this the simplification of the development of complex analysis that uses the steps which are commonly used in traditional database queries.

III. DISCUSSION

Neural signal processing and EEMD: Neural signal analysis is a process of, "detection, diagnosis, and treatment of brain disorders and the associated diseases" [10]. Neural signals are basically nonlinear and non-stationary by nature. As given in introduction part, mainly there are two directions for the neural processing system. Storage management and parallel neural processing

A. EEMD data storage using hadoop

As given in [11], Neuro imaging data can be handled by combining DBMS with grid computing and cluster computing.

Challenges in traditional method:

- On traditional DBMS system, the dense computational task which involves performance of composite analysis of functional connectivity between brain regions is not possible.
- Complex analysis may include recurrent average of subsets of time series (TS) data and comparing TS data, but outdated frameworks for data storage are less fortified for this task.
- The immediate result of these limitation turned into bottleneck problem that is, "It results in slower data analysis, reduces the number of questions that can be asked of the data, and makes it difficult to enable concurrent access to the data (for local and remote users) as is often needed for complex analyses and collaborative research [12]."

Advantages of using big data against traditional method:

- Different compute nodes with duplicating computing task, has fault-tolerant storage and data handling..
- High throughput data processing gets imparted with the help of batch processing model and a productive file system, known as HDFS.

Process to use big data in neural processing:

Ensemble Empirical Mode Decomposition (EEMD) being both compute intensive and data intensive, has become a profound result in the field of neural signal. In data intensive programming, the implementation of Hadoop and the MapReduce computing paradigm, has been consider as widely accepted programming model. The epoch level parallelism and trial level parallelism are the multiple parallelisms in the part of EEMD neural signal processing. A active Hadoop cluster on the Future Grid test bed, is an advanced cyber infrastructure with implementation of parallel EEMD processing. Justification of the design and implementation had been done with the help of test results. The aim for the future is to upload large EEG datasets to HDFS with highly efficient runtime support and to overcome multiple level parallelisms by implementing EEMD on large Hadoop cluster [8].

B. Using Apache spark and GPU Processing for fMRI data processing

Even though after the exposure of Big Data techniques for analysing in the field of fMRI, they are not yet frequently engaged in evaluating of neuroimaging data. Big Data technologies also provide and ideal environment for their application. Inadequate interfaces to data formats of neuroscience with big data technologies had made it complex for researches in neuroscience[9].

At this instance Scala is the only preferable language for the development of the Spark framework. For fMRI data analysis different languages are available for interfaces like python and R. Even though having API from one of those languages and with the help of Spark framework, it still cannot deliver admittance to the complete collection of analysis tools accessible in the Scala API[9].

It was said by author, Boubela Roland N., Kalcher Klaudius, Huf Wolfgang, Našel Christian, Moser Ewald in the title, Big Data Approaches for the Analysis of Large-Scale fMRI Data Using Apache Spark and GPU Processing that "Transferring fMRI computations into a Big Data analysis framework like Spark offers the advantage of the direct availability of arrange of tools optimized for particular problems. Two of the most notable application here are machine learning and graph data analysis, provided by the Spark libraries *MLlib* and *GraphX*, respectively. Both machine learning and graph analysis are rapidly growing sub fields in the fMRI community, but the applications of these methods is often limited by the computational means available for tackling the comparatively complex calculations involved [9]."

They also said that "Efficiency in the sense of computation speed as well as efficiency in terms of development time is important in practical research software development. Parallelization tools are available in multiple programming languages at different levels, one of the advantages of Spark in this respect is the relative ease with which it allows for distributing computations in cluster environments even in an interactive shell. As shown in code listing 4, the details of the distribution of computations is hidden from the developer, allowing for easier programming compared to other tools requiring explicit parallization. Furthermore, ease of access could be further improved by convergence with open data pipelines as developed in the context of data sharing

initiatives, as the inclusion of big data tools into published analysis pipelines could help spread such tools to a wider community of researchers who might otherwise not investigate these opportunities"[9].

As Freeman et al. (2014) have shown in their work that "using large amount of quickly available cloud computing resources can conveniently be leveraged using the Spark Framework. For example, in addition to running the Spark Framework, the Amazon web services (AWS) cloud (as used by Freeman et al., 2014) also provides compute nodes with GPUs (https://aws.amazon.com/ec2/instance-types/), and therefore, could also be employed for the GPU accelerated computation of connectivity graphs as proposed herein [9]."

C. Processing of Electrophysiological Signal Data in Neuroscience Applications Using Apache Pig

Sahoo SS, Wei A, Valdez J, Wang L, Zonjy B, Tatsuoka C, Loparo KA and Lhatoo SD in their article for NeuroPigPen: A Scalable Toolkit for Processing Electrophysiological Signal Data in Neuroscience Applications Using Apache Pig said that, "Apache Pig is a Scalable Toolkit for handling Electrophysiological Signal Data in Neuroscience Applications. [1] In this toolkit the primary advantage of the NeuroPigPen toolkit was compared to hand crafted MapReduce applications. The use of high- level data flow programming constructs defined in Apache Pig without negotiating the performance of the toolkit in terms of scalability. Moreover, the NeuroPigPen has a limitation as its efficiency is not as good as that of hand crafted MapReduce programs since the generic MapReduce tasks generated by the Apache Pig compiler are not optimized for processing electrophysiological signal data". Here in this article, "a comparison of NeuroPigPen with MapReduce programs has been done for future systematic characterization of the differences in performance. This systematic comparison will benefit the users in making a right decision concerning the applicability of NeuroPigPen for easy integration and availability or development of hand crafted MapReduce programs. As well as, there are other concerns of latency associated with initialization of Hadoop MapReduce applications which disrupt the performance of NeuroPigPen modules. This latency is particularly applicable for smaller datasets. Generally the Applications need to perform well on MapReduce architecture then only it will perform well in NeuroPigPen. This aspect of neuroscience applications should be taken in to consideration before making a decision regarding the use of NeuroPigPen toolkit".

They also mentioned other works like, "The scientific workflow systems, such as Taverna (Hull et al., 2006) and Kepler (Ludäscher et al., 2006), can be integrated with NeuroPigPen modules, which will allow wider implemetation of this for processing large datasets by the neuroscience research community. These Scientific workflows are widely used in the bioinformatics and medical informatics community to automate data processing across distributed computing resources with support for failure recovery and ability to collect provenance metadata for scientific reproducibility. Although, there is a clear synergy between the NeuroPigPen toolkit and scientific workflow systems, we have to address the lack of support for workflow system in NeuroPigPen modules". For example, "there is necessity for development of remotely accessible APIs, such as Representational State Transfer (RESTful) Web services, to allow workflow engines to invoke the NeuroPigPen modules. In addition, there is a need to use CSF as the common data representation format by workflow engines to support exchange of signal data between the workflow systems and neuroscience applications. Therefore, as part of our future work, "we plan to develop RESTful APIs for NeuroPigPen modules. This will helps for scientific workflow systems to control the advantages of Hadoop framework (e.g., large scale) for processing and analysing neuroscience data particularly in the broader context of combined multi-modal recordings". We are in process to develop new NeuroPigPen modules to support imaging data used in neuroscience application, for example, classification of cognitive and neural correlates for solving mathematical problem using fMRI [1]."

TABLE II. VARIOUS TECHNIQUES

Attributes	**Hadoop**	**Apache Spark**	**Apache Pig**	**Hive**
Definition	"Apache Hadoop is an open-source software framework used for distributed storage and processing of dataset of big data using the MapReduce programming model. It consists of computer clusters built from commodity hardware" [13].	"Apache Spark is an open-source engine developed specifically for handling large-scale data processing and analytics. Spark offers the ability to access data in a variety of sources, including Hadoop Distributed File System, OpenStack Swift, Amazon S3 and Cassandra" [14].	"Apache Pig is an open-source technology that offers a high-level mechanism for the parallel programming of MapReduce jobs to be executed on Hadoop clusters" [15].	"Apache Hive is a component of Hortonworks DataPlatform. Hive provides a SQL-like interface to datastored in HDP. In the previous tutorial, we used Pig, which is a scripting language with a focus on dataflows. Hiveprovides a database query interface to Apache Hadoop"

Usefulness in neuroscience	It is used to store the neurons data in hadoop cluster in fault tolerant manner. High throughput data processing gets imparted with the help ofbatch processing model and a productive file system, known as HDFS.	For specified problems, shifting fMRI computations into Spark deals the benefit of the uninterrupted accessibility of tools. The spark libraries,MLlib and GraphX provide the two utmost prominent applications, which are machine learning and graph data analysis.	The key benefit of the Pig compared to traditional MapReduce applications is the usage of high- level data flow programming constructs distinct in Apache Pig without compromising the performance of the toolkit in terms of scalability.	The brighter side is that it reassembles the SQL, so querying the neuroscience data becomes easy with the help of Hive.
Best for	Storing	Processing	Programming	Query

IV. Conclusion and Future Scope

After studying various techniques of big data it can be concluded that different techniques have their own existence in particular method. For example Hadoop can be used to store the huge amount of neurological data. Apache spark can be useful to speed up the processing up to 100 percent faster than map reduce. It can be used to analyses the data. While Apache pig can be used to program the data. .For example NeuroPigPen is the result of using pig as a programming language. If more numbers of queries should generated, than Hive is best as it reassembles traditional structured query language. It can be concluded that in current scenario NeuroPigPen is the best toolkit for classification of seizure networks in epilepsy patients and computing functional connectivity network. So, the selection of big data method highly depends on it s application.

References

[1] Sahoo SS, Wei A, Valdez J, Wang L, Zonjy B, Tatsuoka C, Loparo KA and Lhatoo SD (2016) NeuroPigPen: A Scalable Toolkit for Processing Electrophysiological Signal Data in Neuroscience Applications Using Apache Pig. Front. Neuroinform. 10:18. doi: 10.3389

[2] https://en.wikipedia.org/wiki/Neuron

[3] http://www.human-memory.net/brain_neurons.html

[4] Anupama.H.S , N.K.Cauvery , Lingaraju.G.M, Brain Computer Interface And Its Types - A Study, International Journal of Advances in Engineering & Technology, May 2012

[5] Sarah N. Abdulkader, Ayman Atia, Mostafa-Sami M. Mostafa, Brain computer interfacing: Applications and challenges, In Egyptian Informatics Journal, Volume 16, Issue 2, 2015, Pages 213-230, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2015.06.002.

[6] https://en.wikipedia.org/wiki/Big_data

[7] http://blog.cloudera.com/blog/2012/07/processing-rat-brain-neuronal-signals-using-a-hadoop-computing-cluster-part-i/

[8] Lizhe Wang, Dan Chen, Rajiv Ranjan, Samee U. Khan, Joanna Kołodziej, and Jun Wang, Parallel Processing of Massive EEG Data with MapReduce at IEEE 18th International Conference on Parallel and Distributed Systems 2012.

[9] Boubela RN, Kalcher K, Huf W, Našel C and Moser E (2016) Big Data Approaches for the Analysis of Large-Scale fMRI Data Using Apache Spark and GPU Processing: A Demonstration on Resting-State fMRI Data from the Human Connectome Project. Front. Neurosci. 9:492. doi: 10.3389/fnins.2015.00492.

[10] Dennis McFarland, A. Lefkowicz, and Jonathan Wolpaw. Design and operation of an eeg-based brain-computer interface with digital signal processing technology. Behavior Research Methods, 29:337 – 345, 1997.

[11] Uri Hasson, Jeremy I Skipper, Michael J Wilde, Howard C Nusbaum, and Steven L Small. Improving the analysis, storage and sharing of neuroimaging data using relational databases and distributed computing. NeuroImage, 39(2):693–706, 2008.

[12] Hasson U, Skipper JI, Wilde MJ, Nusbaum HC, Small SL. Improving the Analysis, Storage and Sharing of Neuroimaging Data using Relational Databases and Distributed Computing. NeuroImage. 2008;39(2):693-706. doi:10.1016/j.neuroimage.2007.09.021.

[13] https://en.wikipedia.org/wiki/Apache_Hadoop

[14] https://www.webopedia.com/TERM/A/apache-spark.html

[15] http://searchdatamanagement.techtarget.com/definition/Apache-Pig

Copy Right © INDIACom-2018; ISSN 0973-7529; ISBN 978-93-80544-28-1