I J C T A, 9(40) 2016, pp. 309-318

Big-data: Technology Application Challenges and Opportunities

Sagar S. Jambhorkar[*] and Vijay S. Jondhale[**]

ABSTRACT

We are toward the start of a big data period when information is created at an incredible speed from all around. Computing has turned out to be worldwide, number of gadgets like mobile phones, PDAs, portable PCs; individual sensors are making incalculable new advanced seas of data. BigData is a term for a gathering of information sets so vast and complex that it gets to be distinctly hard to process utilizing close by database administration instruments or conventional information preparing applications. Overseeing and preparing expansive information sets is troublesome with existing customary database frameworks. Hadoop and Map Reduce has turned out to be a standout amongst the most intense and prominent instruments for huge information handling. Hadoop and Map Reduce a capable programming model is utilized for breaking down extensive arrangement of information with parallelization, adaptation to internal failure and load adjusting and different components are it is versatile, versatile, and proficient. The difficulties incorporate catch, stockpiling, look, sharing, exchange, examination and representation of Big Data. In this paper the researcher focused on big-data: technology application challenges and opportunities in current scenario.

Keywords: BigData, Hadoop, Map Reduce

1. INTRODUCTION

Big Data are on a very basic level unique in relation to customary measurable investigation on little examples. Big Data is frequently boisterous, alterable, heterogeneous, interrelated and conniving. By the by, even boisterous Big Data could be more important than small examples since general insights got from regular examples and connection investigation for the most part overwhelm singular vacillations and frequently unveil more solid concealed examples and learning. Further, interconnected Big Data shapes expansive heterogeneous data systems, with which data excess can be investigated to make up for missing information, to crosscheck clashing cases, to approve reliable connections, to reveal intrinsic groups, and to uncover shrouded connections and models.

The estimation of Big Data examination in social insurance, to take only one case application space, must be acknowledged on the off chance that it can be connected heartily under these troublesome conditions. On the other side, learning created from information can help in adjusting mistakes and evacuating vagueness. Big Data is additionally empowering the up and coming era of intelligent information examination with real-time answers. Later on, inquiries towards Big Data will be naturally produced for substance creation on sites, to populate hotlists or proposals, and to give a specially appointed investigation of the estimation of information set to choose whether to store or to dispose of it. Scaling complex question preparing systems to terabytes while empowering intelligent reaction times is a noteworthy open research issue today.

Big Data examination is the absence of coordination between database frameworks, which have the information and give SQL questioning, with investigation bundles that perform different types of non-SQL handling, for example, information mining and measurable investigations. Today's examiners are blocked

[*] Department of Computer Science, National Defence Academy, Pune Maharashtra, *E-mail: sjambhorkar@yahoo.co.in*

[**] Research Student, Singhania University, Jhunjhunu Rajasthan, *E-mail: Vijay.jondhale@gmail.com*

by a repetitive procedure of sending out information from the database, playing out a non-SQL process and bringing the information back. This is a snag to persisting the intuitive style of the original of SQL-driven OLAP frameworks into the information mining sort of examination that is in expanding request. A tight coupling between decisive inquiry dialects and the elements of such bundles will profit both expressiveness and execution of the investigation.

Hadoop MapReduce is the most famous innovation of enormous information. A Hadoop MapReduce for the most part comprises of two client deûned capacities: delineate lessen. The contribution of a Hadoop MapReduce employment is an arrangement of key-esteem sets (k; v) and the guide capacity is required each of these sets. The guide work produces at least zero middle of the road key-esteem sets. At that point, the Hadoop MapReduce structure gathers these transitional key-esteem matches by halfway key k and calls the lessen work for every gathering. At long last, the decrease work produces at least zero accumulated outcomes. The excellence of Hadoop MapReduce is that clients generally just need to deûne the guide and diminish capacities. The system deals with everything else, for example, parallelisation and failover, The Hadoop MapReduce structure uses an appropriated ûle framework to peruse and compose its information. Normally, Hadoop MapReduce utilizes the Hadoop Distributed File System (HDFS), which is the open source partner of the Google File System. In this manner, the I/O execution of a Hadoop MapReduce work emphatically relies on upon HDFS. In the ûrst piece of this instructional exercise, we will present Hadoop MapReduce and HDFS in detail. We will balance both with parallel databases. Specifically, we will appear and clarify the static physical execution plan of Hadoop MapReduce and how it influences work execution.

Categorization of big data

The 3Vs categorization of big data is Volume, Variety and Velocity.

Volume

Huge information infers gigantic volumes of information. It used to be workers made information. Since information is created by machines, systems and human cooperation on frameworks like online networking the volume of information to be dissected is monstrous.

Variety

Assortment alludes to the many sources and sorts of information both organized and unstructured. We used to store information from sources like spreadsheets and databases. Presently information comes as messages, photographs, recordings, checking gadgets, PDFs, sound, and so on. This assortment of unstructured information makes issues for capacity, mining and breaking down information.

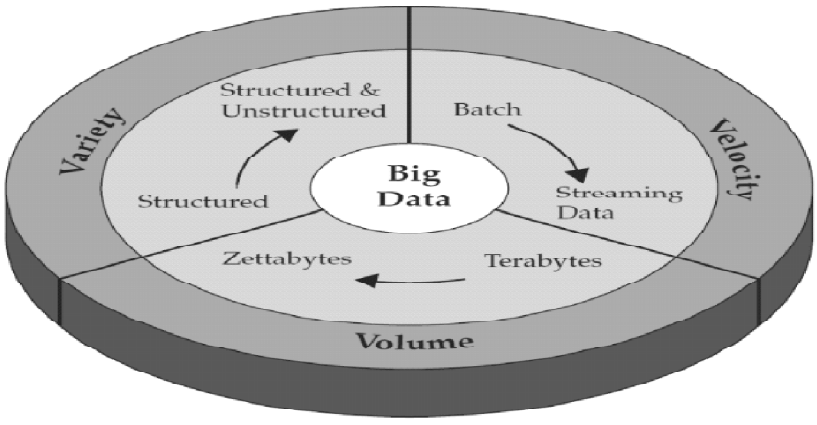

Figure 1: Categorization of big data

Velocity

Big Data Velocity manages the pace at which information streams in from sources like business procedures, machines, systems and human connection with things like online networking locales, cell phones, and so on. The stream of information is huge and constant.

2. REVIEW OF LITERATURES

Protection is the most delicate issue, with applied, lawful, and innovative ramifications. In its thin sense, security is characterized by the International Telecommunications Union as the "privilege of people to control or impact what data identified with them might be revealed." Privacy can likewise be comprehended in a more extensive sense as incorporating that of organizations wishing to ensure their aggressiveness and buyers and states energetic to protect their power and subjects. In both these elucidations, protection is a general worry that has an extensive variety of suggestions for anybody wishing to investigate the utilization of Big Data for advancement—versus information obtaining, capacity, maintenance, utilize and presentation. Security is a crucial human right that has both inherent and instrumental qualities. Two creators, Helbing and Balietti[16], push the need to guarantee a proper level of security for people, organizations and social orders on the loose. In their words, "a current society needs [privacy] keeping in mind the end goal to thrive." Without protection, security, differing qualities, pluralism, development, our fundamental flexibilities are at hazard. Vitally, these dangers concern even people who have "nothing to cover up." There is no compelling reason to grow finally on the significance and affectability of data for enterprises and states. Concentrating on individual security, it is likely that, much of the time, the essential makers—i.e. the clients of administrations and gadgets creating information—are ignorant that they are doing as such, and additionally what it can be utilized for. For instance, individuals routinely agree to the gathering and utilization of web-created information by just ticking a case without completely acknowledging how their information may be utilized or misused [17]. It is likewise vague whether bloggers and Twitter clients, for example, really agree to their information being analysed [18]. what's more, late research demonstrating that it was conceivable to 'de-anonymise' beforehand anonymised datasets raises concerns. The abundance of individual-level data that Google, Facebook, and a couple of cell phone and Visa organizations would together hold on the off chance that they ever were to pool their data is in itself concerning. Since protection is a mainstay of majority rule government, we should stay caution to the likelihood that it may be traded off by the ascent of new advances, and set up every single essential defend.

Get to and sharing although a significant part of the freely accessible online (information from the "open web") has potential esteem for advancement, there is significantly more important information that is firmly held by organizations and is not available for the reasons depicted in this paper. One test is the hesitance of privately owned businesses and different organizations to share information about their customers and clients, and in addition about their own operations. Impediments may incorporate legitimate or reputational contemplations, a need to secure their aggressiveness, a culture of mystery, and, all the more extensively, the nonappearance of the correct motivating force and data structures. There are additionally institutional and specialized difficulties—when information is put away in spots and ways that make it hard to be gotten to, exchanged, and so on. (For instance, MIT teacher Nathan Eagle regularly episodically portrays how he invested weeks in the storm cellars of cell phone organizations in Africa seeking through several containers topped with attractive back-off tapes to accumulate information. An Indonesian portable bearer assessed that it would take up to a large portion of a day of work to concentrate one day of reinforcement information at present put away on attractive tapes.50) Even inside the UN framework it can demonstrate hard to motivate offices to share their program information, for a blend of a few or all of reasons recorded previously. Connecting with suitable accomplices in general society and private divisions to get to non-open information involves setting up non-insignificant legitimate game plans keeping in mind the end goal to secure

(1) Reliable access to information streams and

(2) Get access to go down information for review investigation and information preparing purposes?

There are other specialized difficulties of between equivalence of information and between operability of frameworks, however these may be generally less hazardous to manage than getting formal get to or concurrence on authorizing issues around information. For Big Data for Development to pick up footing, these are not kidding, represent the deciding moment challenges. Any activity in the field should completely perceive the remarkable quality of the security issues and the significance of taking care of information in ways that guarantee that protection is not bargained. These worries must sustain and shape on-going level headed discussions around information protection in the computerized age in a useful way with a specific end goal to devise solid standards and strict guidelines—supported by sufficient devices and frameworks—to guarantee "security safeguarding investigation." in the meantime, the guarantee won't be satisfied if establishments—fundamentally private companies—decline to share information by and large. In light of these necessities, Global Pulse, for example, is advancing the idea of "information philanthropy,"52 whereby "partnerships [would] step up with regards to anonymize (strip out all individual data) their information sets and give this information to social trailblazers to dig the information for bits of knowledge, examples and patterns in real-time or close realtime."

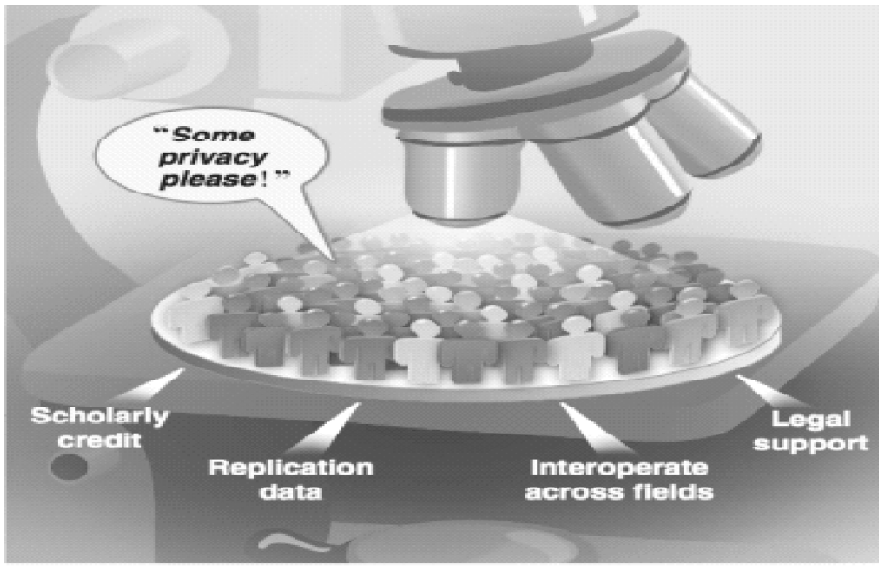

Figure 2: The future of social science data

Source: [19]

Whether the idea of information magnanimity grabs hold or not, it surely indicates the difficulties and roads for thought later on (Figure 3), and we can hope to see facilitate refinements and option models proposed for how to manage protection, and information share.

3. BIG-DATA TECHNOLOGY: SENSE, COLLECT, STORE, AND ANALYZE

The rising significance of Big Data registering originates from advances in a wide range of advances:

Sensors: Digital information are being produced by various sources, including computerized imagers (telescopes, camcorders, MRI machines), compound and natural sensors (microarrays, ecological screens), and even the a large number of people and associations creating site pages.

Computer networks: Data from a wide range of sources can be gathered into huge information sets by means of confined sensor systems, and in addition the Internet.

Data stockpiling: Advances in attractive circle innovation have drastically diminished the cost of putting away information. For instance, a one-terabyte plate drive, holding one trillion bytes of information, expenses around $100. As a source of perspective, it is evaluated that if the greater part of the content in the greater part of the books in the Library of Congress could be changed over to advanced frame; it would indicate just around 20 terabytes.

Cluster computer systems: another type of PC frameworks, comprising of a huge number of "hubs," each having a few processors and circles, associated by fast neighbourhood, has turned into the picked equipment design for information serious registering frameworks. These bunches give both the capacity ability to expansive information sets, and the processing energy to sort out the information, to dissect it, and to react to questions about the information from remote clients. Contrasted and customary elite processing (e.g., supercomputers), where the emphasis is on amplifying the crude Computing force of a framework, bunch PCs are intended to augment the dependability and effectiveness with which they can oversee and break down vast information sets. The "trap" is in the product calculations – group PC frameworks are made out of colossal quantities of modest ware equipment parts, with adaptability, dependability, and programmability accomplished by new programming ideal models.

Distributed Computing offices: The ascent of extensive server farms and bunch PCs has made another plan of action, where organizations and people can lease stockpiling and computing limit, instead of making the substantial capital speculations expected to develop and arrangement expansive scale PC establishments. For instance, Amazon Web Services (AWS) gives both system open stockpiling evaluated by the gigabyte-month and Computing cycles valued by the CPU-hour. Similarly as couple of associations work their own energy plants, we can anticipate a time where information stockpiling and processing get to be utilities that are universally accessible.

Data analysis algorithms: The gigantic volumes of information require mechanized or semi robotized examination – systems to recognize designs, distinguish abnormalities, and concentrate Knowledge. Once more, the "trap" is in the product calculations - new types of calculation, Combining measurable examination, enhancement, and counterfeit consciousness, can develop factual models from substantial accumulations of information and to induce how the framework ought to react to new information. For instance Netflix utilizes machine learning as a part of its proposal framework, anticipating the premiums of a client by contrasting her motion picture seeing history with a measurable model produced from the aggregate review propensities for a huge number of different clients.

4. TECHNOLOGY AND APPLICATION CHALLENGES

A great part of the innovation required for huge information processing is creating at an agreeable rate because of market strengths and mechanical advancement. For instance, circle drive limit is expanding and costs are dropping because of the continuous advance of attractive stockpiling innovation and the extensive economies of scale gave by both PCs and huge server farms. Different angles require more engaged consideration, including:

High-speed networking: Although one terabyte can be put away on plate for just $100, exchanging that much information requires a hour or more inside a bunch and about a day over a run of the mill "rapid" Internet association. (Inquisitively, the most reasonable technique for exchanging mass information starting with one site then onto the next is to deliver a circle drive through Federal Express.)

These transmission capacity restrictions increment the test of making effective utilization of the processing and capacity assets in a group. They likewise constrain the capacity to connect geologically scattered groups and to exchange information between a bunch and an end client. This dissimilarity between the measures of information that is useful to store, versus the sum that is down to earth to convey keep on increasing. We require a "Moore's Law" innovation for systems administration,

where declining costs for systems administration foundation consolidate with expanding data transfer capacity.

Cluster computer programming: Programming extensive scale, circulated PC frameworks is a longstanding test that gets to be distinctly basic to prepare vast information sets in sensible measures of time. The product must disperse the information and calculation over the hubs in a bunch, and recognize and remediate the unavoidable equipment and programming mistakes that happen in frameworks of this scale. Significant developments have been made in strategies to sort out and program such frameworks, including the MapReduce programming structure presented by Google. Considerably more effective and general methods must be created to completely understand the force of huge information processing over numerous spaces.

Extending the reach of cloud computing: Although Amazon is earning substantial sums of money with AWS, innovative confinements, particularly correspondence transmission capacity, make AWS unacceptable for assignments that require broad calculation over a lot of information. In expansion, the transfer speed confinements of getting information all through a cloud office cause extensive time and cost. In a perfect world, the cloud frameworks ought to be geologically scattered to diminish their helplessness because of seismic tremors and different calamities. Yet, this requires much more noteworthy levels of interoperability and information portability. The OpenCirrus venture is pointed in this heading, setting up a universal testbed to permit investigates interlinked group frameworks. On the authoritative side, associations must conform to another costing model. For instance, government contracts to colleges don't charge overhead for capital expenses (e.g., purchasing a substantial machine) however they accomplish for working expenses (e.g., leasing from AWS). After some time, we can imagine a whole biology of cloud offices, some giving non specific Computing abilities and others focused toward particular administrations or holding particular information sets.

Machine learning and other data analysis techniques: As a logical teach, machine learning is still in its initial phases of advancement. Numerous calculations don't scale past information sets of a couple of million components or can't endure the measurable clamor and holes found in true information. Additionally research is required to create calculations that apply in certifiable circumstances and on information sets of trillions of components. The robotized or semi computerized examination of colossal volumes of information lies at the heart of Big Data Computing for all application areas.

Widespread deployment: As of not long ago, the primary trend-setters in this space have been organizations with Internet-empowered organizations, for example, web indexes, online retailers, and person to person communication destinations. Just now are technologists in different associations (counting colleges) getting comfortable with the abilities and devices. Albeit numerous associations are gathering a lot of information, just a modest bunch are making full utilization of the bits of knowledge that this information can give. We expect "huge information science" – frequently alluded to as eScience – to be inescapable, with far more extensive reach and effect even than past era computational science.

Security and protection: Data sets comprising of so much, potentially touchy information, and the apparatuses to concentrate and make utilization of this data offer ascent to numerous conceivable outcomes for unapproved get to and utilize. A lot of our protection of security in the public arena depends on current wasteful aspects. For instance, individuals are checked by camcorders in numerous areas – ATMs, accommodation stores, airplane terminal security lines, and urban convergences. Once these sources are arranged together, and refined Computing innovation makes it conceivable to relate and examine these information streams, the prospect for mishandle gets to be distinctly critical. Likewise, cloud offices turn into a savvy stage for vindictive specialists, e.g., to dispatch a botnet or to apply enormous parallelism to break a cryptosystem. Alongside building up this innovation to empower valuable capacities, we should make shields to anticipate mishandle.

5. CHALLENGES AND OPPORTUNITIES WITH BIG DATA

We are inundated with a surge of information today. In an expansive scope of use regions, information is being gathered at phenomenal scale. Choices that beforehand depended on mystery, or on carefully developed models of reality, can now be made in light of the information itself. Such Big Data investigation now drives about each part of our current society, including versatile administrations, retail, producing, budgetary administrations, life sciences, and physical sciences. Logical research has been reformed by Big Data [1]. The Sloan Digital Sky Survey [2] has today turned into a focal asset for stargazers the world over. The field of Astronomy is being changed from one where taking photos of the sky was a huge part of a stargazer's business to one where the photos are all in a database as of now and the space expert's errand is to discover fascinating articles and wonders in the database. In the organic sciences, there is presently a settled custom of keeping logical information into an open storehouse, furthermore of making open databases for use by different researchers. Truth be told, there is a whole teach of bioinformatics that is to a great extent gave to the curation and investigation of such information. As innovation advances, especially with the approach of Next Generation Sequencing, the size and number of trial information sets accessible is expanding exponentially. Big Data can possibly alter look into, as well as training [3]. A late itemized quantitative examination of various methodologies taken by 35 sanction schools in NYC has found that one of the main five arrangements associated with quantifiable scholastic viability was the utilization of information to guide direction [4]. Envision a world in which we have entry to a gigantic database where we gather each itemized measure of each understudy's scholastic execution. This information could be utilized to plan the best ways to deal with training, beginning from perusing, composing, and math, to cutting edge, school level, courses. We are a long way from having admittance to such information, yet there are intense patterns in this course. Specifically, there is a solid pattern for gigantic Web organization of instructive exercises, and creates an inexorably vast measure of itemized information about understudies' execution. It is broadly trusted that the utilization of data innovation can lessen the cost of medicinal services while enhancing its quality [5], by making care more preventive and customized and constructing it in light of more broad (locally established) persistent observing. McKinsey gauges [6] a reserve funds of 300 billion dollars consistently in the only us. In a comparable vein, there have been convincing cases made for the estimation of Big Data for urban arranging (through combination of high-constancy geological information), clever transportation (through investigation and perception of live and definite street organize information), natural displaying (through sensor systems universally gathering information) [7], vitality sparing (through uncovering examples of utilization), keen materials (through the new materials genome activity [6]), computational sociologies (another strategy quickly developing in fame on account of the significantly brought down cost of getting information) [8], monetary systemic hazard examination (through incorporated examination of a web of agreements to discover conditions between money related substances) [9], country security (through investigation of interpersonal organizations and budgetary exchanges of conceivable psychological oppressors), PC security (through examination of logged data and different occasions, known as Security Information and Event Management (SIEM)), et cetera. In 2010, endeavors and clients put away more than 13 Exabyte's of new information; this is more than 50,000 circumstances the information in the Library of Congress. The potential estimation of worldwide individual area information is assessed to be $700 billion to end clients, and it can bring about an up to half abatement in item advancement and get together expenses, as per a late McKinsey report [6]. McKinsey predicts a similarly incredible impact of Big Data in business, where 140,000-190,000 specialists with "profound investigative". Of course, the late PCAST provide details regarding Networking and IT R&D [10] recognized Big Data as a "research outskirts" that can "quicken advance over an expansive scope of needs." Even prevalent news media now acknowledges the estimation of Big Data as confirm by scope in the Economist [11], the New York Times [13], and National Public Radio [12]. While the potential advantages of Big Data are genuine and critical, and some underlying triumphs have as of now been accomplished, (for example, the Sloan Digital Sky Survey), there stay numerous specialized difficulties that must be tended to completely understand this potential. The

sheer size of the information, obviously, is a noteworthy test, and is the one that is most effectively perceived. Notwithstanding, there are others. Industry examination organizations get a kick out of the chance to call attention to that there are difficulties in Volume, as well as in Variety and Velocity [14], and that organizations ought not concentrate on simply the first of these. By Variety, they normally mean heterogeneity of information sorts, representation, and semantic translation. By Velocity, they mean both the rate at which information arrive and the time in which it must be followed up on. While these three are vital, this short rundown neglects to incorporate extra vital prerequisites, for example, protection and ease of use. The investigation of Big Data includes numerous unmistakable stages as appeared in the figure underneath, each of which presents challenges. Many individuals tragically concentrate just on the examination/demonstrating stage: while that stage is essential, it is of little use without alternate periods of the information investigation pipeline. Indeed, even in the investigation stage, which has gotten much consideration, there are inadequately comprehended complexities with regards to multi-rented bunches where a few clients' projects run simultaneously. Numerous noteworthy difficulties stretch out past the examination stage. For instance, Big Data must be overseen in setting, which might be boisterous, heterogeneous and exclude a forthright model. Doing as such raises the need to track provenance and to handle vulnerability and blunder: themes that are vital to achievement, but then once in a while said at the same time as Big Data. Correspondingly, the inquiries to the information investigation pipeline regularly not all are laid out ahead of time. We may need to make sense of good inquiries in light of the information. Doing this will require more brilliant frameworks furthermore better support for client cooperation with the investigation pipeline. Actually, we right now have a noteworthy bottleneck in the quantity of individuals enabled to make inquiries of the information and investigate it [15]. We can definitely expand this number by supporting many levels of engagement with the information, not all requiring profound database aptitude. Answers for issues not originate from incremental changes to the same old thing, for example, industry may make all alone. Or maybe, they oblige us to in a general sense reconsider how we oversee information examination.

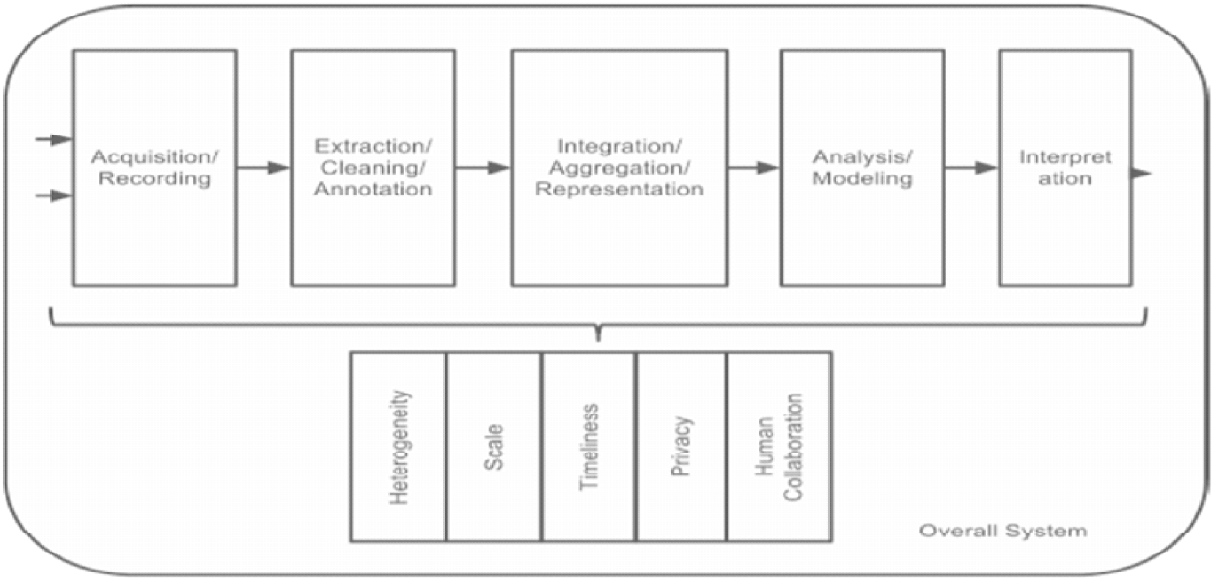

Figure 3: The key steps of big data analysis

Fortunately, existing computational methods can be connected, either as is or with a few expansions, to at any rate a few parts of the Big Data issue. For instance, social databases depend on the thought of coherent information freedom: clients can consider what they need to register, while the framework (with gifted specialists outlining those frameworks) decides how to process it productively. Essentially, the SQL standard and the social information show give a uniform, effective dialect to express many inquiry needs and, on a fundamental level, permits clients to pick between merchants, expanding rivalry. The test in front

of us is to join these sound elements of earlier frameworks as we devise novel answers for the numerous new difficulties of Big Data. In this paper, we consider each of the containers in the figure above, and talk about both what has as of now been done and what challenges stay as we try to misuse Big Data. We start by considering the five phases in the pipeline, then proceed onward to the five cross-cutting difficulties, and end with a talk of the engineering of the general framework that consolidates every one of these capacities.

CONCLUSION

Propels in different branches of technology – data sensing, data communication, data computation, and data storage – are driving a time of exceptional development for data recovery. The world of Big Data is always showing signs of change and delivering colossal measures of information that makes difficulties to prepare the applications utilizing existing arrangements. Huge information applications require processing assets and capacity subsystems that can scale to oversee enormous measures of assorted information, People, organizations, governments, and society overall now have admittance to huge accumulations of huge information, engaging them to construct their own particular examination. Data centers are thusly required to acquaint more hubs with their framework or supplant their current equipment with all the more capable frameworks to react to this developing interest. This pattern builds the framework cost and power utilization. We trust this is the correct time to recognize the correct Computing stage for Big Data investigation preparing that can give a harmony between handling limit and power proficiency.

REFERENCES

[1] Advancing Discovery in Science and Engineering Computing Community Consortium, Spring 2011.

[2] Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra Solar Planetary Systems. Jan-2008, Available at *http://sdss3.org/collaboration/description.pdf*

[3] Advancing Personalized Education, Computing Community Consortium.Spring 2014.

[4] Getting Beneath the Veil of Effective Schools: Evidence from New York City. Will Dobbie, Roland G. Fryer, Jr. NBER. Working Paper No. 17632. Issued Dec. 2011.

[5] Smart Health and Wellbeing. Computing Community Consortium. Spring 2015.

[6] Big data: The next frontier for innovation, competition, and productivity.. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and An gela Hung Byers. McKinsey Global Institute. May 2011.

[7] A Sustainable Future. Computing Community Consortium. Summer 2011.

[8] Computational Social Science. David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, AlbertLaszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gut mann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Science 6 February 2009: 323 (59 15), 721-723.

[9] en, and Louiqa Raschid. Proc. Fifth Biennial Conf. Innovative Data Systems Research, Jan. 2011.

[10] Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology, PCAST Report, Dec.2015 Available at *http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-2010.pdf*

[11] Drowning in numbers – Digital data will flood the planet and help us understand it better. The Economist, Nov 18, 2011 *http://www.economist.com/blogs/dailychart/2011/11/big-data.*

[12] The Search for Analysts to Make Sense of Big Data. Yuki Noguchi, National Public Radio, Nov.30, 2011. *http://npr.org/2011/11/30/14289306/the-search-for-analysts-to-make-sense-of-big-data.*

[13] The Age of Big Data Steve Lohr, New York Times, Feb. 11, 2012 *http://www.nytimes.com/2012/02/12/sunday-review/big-dates-impact-in-the-world.html*

[14] Patten based Strategy, Getting Value from Big Data. Gartner Group press release. July 2011 Available at *http://www.gartner.com/it/page.jsp?id=1731916*

[15] The Age of Big Data. Steve Lohr, New York Times, Feb 11, 2012. *http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html*

[16] Helbing Balietti, "From Social Data Mining to Forecasting Socio-Economic Crisis." The European Physical Journal-Special Topics (Volume 195, Number 1, 3-68, pg 24) 26 July 2011.

[17] Efrti, Amir "Like Button Follows Web Users", The Wall Street Journal 18 May 2011. http://online.wsj.com/articles/SB10001424052748704281504576329441432995616.html

[18] Boyd, Dana and Crawford, Kate "Six Provocations for Big Data", Working Paper Oxford Internet Institute 21 Sept. 2011 *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431*.

[19] King, Gary N. and Eleanor Powell, How Not to Lie Without Statistics Working Paper Harvard University, 22 Aug. 2008 *http://gking.harvard.edu/gking/files/online.pdf*.