

## **BIG DATA “WHAT-HOW-WHY” AND ANALYTICAL TOOLS FOR HYDROINFORMATICS.**

*Dadi Sanyasi Naidu*

*Research Scholar, Department of Geo-Engineering, Andhra University, Visakhapatnam, A.P., India*

[sndadi@gmail.com](mailto:sndadi@gmail.com)



### **ABSTRACT**

*The unavoidable pattern of enormous information alongside the developing ability to deal with colossal datasets is reshaping how we comprehend the world. The International Data Corporation (IDC) report has assessed that the information size of the world will develop from 130 bytes (1018 bytes) in 2005 to 40 zetta bytes (1021 bytes) in 2020, at a 40% yearly increment. New datasets are ceaselessly being gathered from the web, the Internet of Things, the remote detecting system and e-start, wearable gadgets, and so on. Sadly, just 3% of all information is legitimately labeled and prepared to utilize, and just 0.5% of information is breaking down, which yields an expansive potential market for informational usage.*

**Keywords:** Big Date, Hadoop, GIS, Remote Sensing, Hydroinformatics, Database, Data, Cloud Computing.

**Citation:** Dadi Sanyasi Naidu (2018). Big Data “Waht-How-Why” And Analytical Tools For Hydroinformatics. International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR ) ISSN:2581-4281 Vol 1, Issue4, June 2018, #Art.214, pp37-47

### **1. Introduction**

The unavoidable pattern of enormous information alongside the developing ability to deal with colossal datasets is reshaping how we comprehend the world. The International Data Corporation (IDC) report has assessed that the information size of the world will develop from 130 bytes (1018 bytes) in 2005 to 40 zetta bytes (1021 bytes) in 2020, at a 40% yearly increment

[1]. New datasets are ceaselessly being gathered from the web, the Internet of Things, the remote detecting system and e-start, wearable gadgets, and so on. Sadly, just 3% of all information is legitimately labeled and prepared to utilize, and just 0.5% of information is breaking down, which yields an expansive potential market for informational usage [2]. A renowned early endeavor of the enormous information application was Google Flu Trend (GFT) that checked well being looking to conduct as online web seek inquiries by a huge number of clients around the globe consistently. The technique was to locate the best matches among 50



million hunt terms to fit 1152 influenza information focuses from Central Disease Control. GFT assessed the level of week by week flu movement with a one day announcing slack, considerably shorter than the Central Disease Control with two week revealing slack [3]. GFT anticipated the flu movement reacting considerably speedier than the CDC, yet endured its hazardous execution. In 2009, its poor underestimation of the flu like sickness in the United States of the swine influenza pandemic constrained Google to adjust its calculation. GFT overestimated influenza commonness in 100 out of 108 weeks from 21 August 2011 to 1 September 2013 [4]. In December 2012, it overestimated more than twofold the specialist visits for flu like sickness (ILI) than the Central Disease Control [5]. Google quit distributing influenza slant information and began to pass the information to specific associations to engage their examination in summer 2015 [6]. Another utilization of enormous information is exactness showcasing, i.e. the online film membership rental specialist organization Netflix has its suggestion framework in light of hundreds a great many aggregated unknown motion picture evaluations to enhance the likelihood that the clients lease the motion pictures prescribed by Netflix [7]. In spite of the fact that the prominence of enormous information is connected with its business esteem, we trust that the possibility of huge information can profit the hydroinformatics examine for different reasons. To start with, the enormous information examination energizes the use of numerous datasets from different sources to find the huge pattern. Furthermore, the figuring instruments produced for the enormous information investigation, e.g. parallel registering and dispersed information stockpiling, can help handle the information escalated occupations in the field of hydroinformatics. Thirdly, the novel connection

found by mining different vast datasets can possibly prompt new logical investigation. Aside from the organizations in the web business working intimately with the information from the web, the researchers have gathered considerable measure of information for hydrology, meteorology and earth perception with a history any longer than that of the web. The advancement of the web and the development of open information fundamentally quicken the information sharing and enhance the openness of the documented information. The hydroinformatics group will profit by the dynamic mix of an immense measure of information and the information preparing advances for learning disclosure and administration. Precipitation is one vital piece of the water cycle in hydrology. The amassed precipitation datasets from heterogeneous sources, e.g., rain gages, climate radars, satellite remote detecting and numerical climate models, have achieved several terabytes in a measure, with various attributes, i.e., spatial and transient scope, determination, and vulnerabilities. Information combination is a conceivable strategy to use the gathered datasets to deliver a superior outcome with upgraded determination and limited vulnerability. This paper expects to give per users who are not all that comfortable with enormous information with an opportune audit on its idea and the applicable innovation, beginning from the clarification of the idea of huge information, at that point prologue to the prevalent Apache Hadoop family to deal with expansive measure of information. From that point forward, the significance of huge information with hydroinformatics is clarified in three measurements, the normal measurement, the social measurement and the business measurement, to encourage more scientists in the hydroinformatics group to endeavor novel research in light of huge information.



## **2. Enormous information and the significant innovation**

### **2.1. Influence the idea to clear**

The trendy term of 'Huge Data' is now and again so hot that many individuals endeavor to grasp it in this information rich time without an unmistakable comprehension. The idea of huge information began from the extraordinary substantial datasets that have been gathered, however can't be prepared in the middle of the road slipped from time with conventional information handling techniques. The term 'huge information' is basic yet its significance is equivocal. It is regularly used to depict informer indexes with the amount and many-sided quality past the limit of typical figuring devices to catch, clergyman, oversee, and process with a variable speed [8]. Another clarification of Big Data alludes to growing new bits of knowledge or making new esteems at an extensive scale rather than a little one [9]. A formal meaning of huge, information is the data resources portrayed by such a high volume, speed and assortment to require particular innovation and logical techniques for its change into esteem, in view of examination of 14 existing meanings of enormous information [10]. This definition can be subdivided into three gatherings: the qualities of the informational indexes, the particular advancements and expository strategies to control the information, and the plans to separate bits of knowledge from the information and production of new esteems. In this way, enormous information isn't just about huge measures of information. All in all, the objective of the enormous information investigation is learning, revelation from gigantic informer indexes, which is a testing deliberate issue. The information examination frameworks ought to

use the current equipment stage with conveyed and parallel figuring, suit an assortment of information designs, models, misfortune capacities and strategies, be exceptionally adaptable for clients to indicate their information investigation objectives through an expressive however basic dialect, give valuable representations of key parts of the examination, speak with other computational stages flawlessly, and give a large number of abilities well-known from extensive scale databases [11].

### **2.2. The Map Reduce parallel registering**

The Map Reduce parallel registering is the new figuring model highlighting parallel information preparing to accelerate the information I/O effectiveness [10], created in the enormous information time. The inspiration of such a figuring technique is, to the point that more accentuation has been put on information I/O separated from the registering procedure itself. The worry is whether the current processing framework can deal with the undeniably vast information inside middle of the road time. The information stockpiling limit expanded drastically in the previous decades. In 2014, Western Digital dispatched the 8 TB hard drive and reported the world initial 10 TB hard drives [12]. The unit cost of information stockpiling will drop down from \$2.00 per GB to \$0.20 per GB from 2012 to 2020 [1]. The capacity of information should never again be a major issue, owing to the huge stockpiling innovations, for example, Direct Attached Storage (DAS), Network Attached Storage (NAS) and Storage Area Network (SAN), and in addition the cloud information stockpiling. Be that as it may, the I/O speed of the hard plate becomes gradually constrained by the hard circle instrument. Strong state circle (SSD) has a substantially higher I/O rate and



immaterial look of time, meanwhile, the cost per unit stockpiling is considerably higher than that of the hard plate. Despite the cost, the SSD has a lower stockpiling limit of single gadget. The I/O speed of the information stockpiling gadgets is the bottleneck of outrageous vast information preparing as opposed to the information stockpiling limit.

A fitting, programming, framework is fundamental to managing to a great degree expansive data sets separated from the improvement of the equipment framework. As the change of I/O speed of the equipment framework did not get the speed of the extension of information stockpiling, the time required to process the information significantly expanded without a suitable calculation. The parallel processing and circulated stockpiling were produced to experience this issue. MapReduce is a disseminated programming model for preparing and producing huge datasets created by Google. The possibility of MapReduce is to determine a Map and a Reduce work which are appropriate for parallel figuring, and the basic runtime framework naturally parallelizes the calculation crosswise over huge scale groups of machines, handles machine disappointments [9], and calendars between machine correspondence to make effective utilization of the system and plates. As the extent of datasets is to a great degree substantial for huge information issues, a group of machines associated with a system is utilized to beat the cutoff of figuring force and information stockpiling of a solitary machine, however the system transfer speed turns into the bottleneck as it is an uncommon asset. In this manner, the MapReduce framework is enhanced focusing on decreasing the information exchange over the system through sending the code to the neighboring machine and composing the middle of the road information to

nearby circle. The MapReduce framework limited the effect of moderate machines, and can deal with machine disappointments and information misfortune by repetitive execution [13]. The Hadoop is an open-source variant of the MapReduce structure created by Apache, unreservedly accessible to an academic group. The Hadoop contains the Hadoop Distributed File System (HDFS) cooperating with MapReduce after Google distributed the specialized points of interest of the Google File System [14], aside from which the Apache Hadoop likewise contains Hadoop Common, the basic utilities that help the other Hadoop modules; and Hadoop YARN, a structure for work booking and group asset administration. There are numerous different activities in Apache which are identified with Hadoop, including HBase (an adaptable, appropriated database that backings organized information stockpiling for expansive tables), Hive (an information stockroom foundation that gives information synopsis and specially appointed questioning), Mahout (a versatile machine learning and information mining library), Pig (an abnormal state information stream dialect and execution system for parallel calculation) and ZooKeeper (an elite coordination benefit for disseminated applications), etc.[15]. Hadoop MapReduce has a shortcoming amid iterative information examination that the irregular datasets are put away on the neighborhood hard circle. As the iterative information investigation requires numerous read and compose of nearby irregular information, which will significantly back off the examination. This happens to most machine learning calculations, e.g., inclination better than average. Apache Spark is the most recent programming model in the enormous information world, highlighting its exceptionally quick information preparing speed for iterative employments [16]. The Spark accomplished its



exceptionally quick speed by the executing Resilient Distributed Datasets (RDDs), a circulated memory, reflection that gives the software engineers a chance to perform in-memory calculation [17]. The Spark beats Hadoop by 20 times in speed by using the RAM rather than hard plate to store the irregular information.

### Significance to hydroinformatics

Hydroinformatics, began from the computational power through pressure, includes the utilization of data and correspondences innovations (ICTs) to the comprehension and administration of the waters of the world [18], tending to the undeniably major issues of the fair and effective utilization of water for various purposes. Once the term hydroinformatics was characterized, it intended to incorporate counterfeit consciousness to the numerical recreation and demonstrating, and to move the computational-serious examination of data based research. The two principle lines of hydroinformatics, information digging for learning disclosure and learning administration [19], are unequivocally subject to data of which information, both literary or non-printed, is the significant bearer. Information from savvy meters, shrewd sensors and keen administrations, remote detecting, earth perception frameworks, and so forth., will provoke hydroinformatics into the unavoidable huge information time. The test of enormous information and information digging for natural tasks is the most squeezing one sooner rather than later [20]. All in all, the water-related issues are very perplexing because of the hiding interrelationships between water-related natural, social and business factors. The information being produced and gathered applicable to hydroinformatics highlights enormous volumes and numerous sorts. With the end

goal of improving, the information hotspots for the hydroinformatics, without loss of liberality, can be ordered into three measurements, i.e., the regular measurement, the social measurement, and the business measurement.

### 3.1. The common measurement

The most common measurement is about water as one critical segment of the indigenous habitat. Understanding the water cycle, the transient and spatial dispersion of water and the connection of water and the earth is a piece of the targets of hydroinformatics for enhancing the water asset administration, surge and dry season administration. The water-related information incorporates the estimations of (precipitation, snow and hail), waterway stream, water quality, soil dampness, soil trademark, ground water condition, air temperature and moistness, sun powered transition, and so on. The perception techniques created from nearby station for guide estimation toward remote detecting - radar and satellites, and automaton. The earth perception satellites are creating immense volume of information, including climate and water-related data. ESA has propelled SMOS for soil dampness perception in 2009, and will dispatch ADM-Aeolus for Atmospheric Dynamics perception in 2017 [21]. NASA propelled SMAP to delineate dampness and decide the stop or defrost state in 2015 [22]. The GPM mission propelled in 2015 intends to give worldwide rain and snow perception in light of the accomplishment of TRMM propelled in 1997 [23]. EUMETSAT has two ages of dynamic METEOSAT satellites in geostationary circle and a progression of three polar circling METOP satellites for climate now casting and anticipating and understanding the environmental change. Without question, the



expanding measure of earth perception information, including precipitation, soil dampness and wind speed and so on., will enhance the comprehension of the worldwide water cycle, and advantage the climate estimating, surge and dry spell expectation. Tragically, albeit many satellites were propelled or to be propelled, the enormous measure of accessible information is once in a while utilized, just three to five percent of information is utilized on day by day normal, while billions of dollars have been contributed yearly [24]. Aside from the earth perception information, reanalysis information is another vital data source with high information quality. At the end of the day, the data source isn't restricted to the perception of the present circumstance and the documented past circumstance, the model produced information can't be dismissed. Reanalysis of chronicled perceptions is accomplished by consolidating propelled gauge models and information absorption frameworks to make worldwide informational indexes of the environment, arrive surface, and seas, as an operational examination dataset will experience the ill effects of irregularity because of the continuous upgrades of the figure models. The NCEP Climate Forecast System Reanalysis incorporates more than 80 factors, backpedals to 1948 and is proceeding with [25]. ECMWF has arrangement of ERA ventures for worldwide air reanalysis following back to 1957 [26]. The Japan Meteorological Agency directed the JRA-55 venture for a superb homogeneous atmosphere dataset covering the last 50 years [27]. The model produced information is four dimensional, three measurements in space and one in time, and of high spatial and fleeting scope and determination, bringing about enormous volume of information, which implies the hydroinformatics is entering an information concentrated time. Usage of the as of now accessible

information is testing couple to the vulnerabilities of the information, the difficulties of preparing and the absence of thoughts on information use. In the huge information period, it is urged to make the best of the tremendous measure of information with the resilience of the vulnerabilities. The handling of huge measure of datasets is getting to be noticeably less demanding with the advancement of registering instruments. The absence of inventive thoughts is the principle confinement of the use of information. A boondocks application illustration is a model programming that consequently discovers a perfect area for hydro-power based on over 30 freely remote sensing and environmental datasets in UK [28].

### **3.2. The social measurement**

The social measurement is about the association of water condition and the human culture. With the digitization of printed data accessible on the web and the blast of online networking literary mining advances empower the new research territory of general society disposition towards certain issues. For example, 5 million logical articles broke down to investigate the effect of the Fukushima fiasco on the media state of mind towards atomic power [29]. Comparable thoughts can be moved to find water-related issues, e.g., the social disposition towards environmental change, water sparing, water arrangement, and so forth. Aside from the revelation of open state of mind, the web is logging the exercises of web clients, which can be conceivably profitable for finding true circumstances showed by the case of Google Flu Trend specified in the past area. The Twitter information is presently pulling in numerous specialists to dive into for water condition related research. It was discovered that Twitter substance could deduce every day precipitation rates in five UK urban communities,



which uncovered the online literary highlights in Twitter were emphatically identified with the theme with huge induction [30]. Two Dutch associations, Deltares and Floodtags, have built up a continuous surge degree maps in light of tweets about surges in Jakarta, Indonesia [31]. This strategy gives the debacle administration an ongoing perspective of the circumstance with a wide scope. The improvement of the new media information on the web empowers another model for logical research. The new model accumulates data from what the web clients post on the web. The clients are really acting a part of data gathering, and they store the data about what they see about the earth to the web. The web resembles an unlimited sea of information that records how the web clients cooperate with the web. The information sea has a profitable potential for researchers to find novel connections between's genuine circumstances. The central information mining strategies behind the huge information application, for example, Google Flu Trend, assessing precipitation from Twitter, and so forth., is the same, i.e. to uncover the connection between's the data and the focused on results. The qualification of these investigations is that the interpersonal, organization information application depends on individuals' mental response to specific occasions while the nature logical research is fundamental in view of the physically interpretable model. As the conduct of individuals is equivocal to decipher and anticipate, the huge information investigation of interpersonal, organization information is commanded by the machine learning or measurable methodologies.

### **3.3. The business measurement**

The business measurement covers yet not constrained to water extraction, water treatment, water supply, squander water gathering and treatment. IBM has been a pioneer in using information and figuring devices coordinated effort with NOAA to investigate the matter of climate. They assembled one of the principal parallel preparing supercomputers for climate, demonstrating in 1995, named as Deep Thunder Project. Profound Thunder makes 24-to 48-hour estimates at 1 - 2 km determination with a lead time of three hours to three days and consolidates with other information altered for business purposes, for example, to enable a service organization to get ready for the eventual outcomes of a noteworthy tempest or to enable carriers and airplane terminals to deal with the climate created delays by improving or joining flights all the more effectively [32]. Another plausibility is that, as roused by the enormous information application in web based business that use the amassed client movement logs for proposal framework, the keen metering information can be incorporated with end-client water utilization information, remote correspondence systems and data administration frameworks so as to give continuous data on how, when and where water is being expended for the purchaser and utility [33]. The data from the blend of information will be important to engineers, designers and organizers, trying to comprehend water utilization designs for future water arranging. More brilliant metering is one case of the aspiring thought of the Internet of Things as a worldwide foundation for the data society, empowering propelled benefits by interconnecting things in light of existing and advancing interoperable data and correspondence innovations [34]. Moreover, the operation information gathered by



organizations in the water business additionally has potential esteems for information digging for improving the framework and giving more data to basic leadership.

### **3.4. The pattern of open information**

The expanding number of transparently accessible information sources will profit the exploration group as information is the essential material for research based information. Open information implies information that can be unreservedly utilized, adjusted, and shared by anybody for any reason [35]. Open information is the further advancement of free information that information is uninhibitedly authorized for constrained purposes and certain clients, while shut information is typically limited by copyright, licenses or different instruments. The objectives of the open information development are like those of other "Open" developments, for example, open source, open equipment, open substance, and open access. The information proprietor might not have the proper thoughts and strategies to deliver additional esteems from the information, while, then again, individuals with imaginative thoughts and capacity of preparing the information may think that it's hard to discover and get to the information they require. The open information developed will actuate the mix of information, information mining techniques and new plans make extra esteems by evacuating the boundary between the information suppliers and the information clients. In this way, the exploration information and its items can accomplish the full esteem and quicken the future research just while being open. Numerous national governments made sites for the open conveyance of their information for straightforwardness and responsibility, e.g., Data. Gov for the US government, Data. Gov. UK for the UK government,

European Union Open Data Portal (<http://open-data.europa.eu/>) and Canada's Open Government entrance (<http://open.canada.ca/en>) and so forth. For open information in science, the World Data System (WDS) of the International Council for Science was made in view of the inheritance of the World Data Centers in 2008 to guarantee the widespread and impartial access to quality-guaranteed logical information, informational administrations, items and data. National Climatic Data Center, containing immense measure of ecological, meteorological and atmosphere informational collections, is the world's biggest file of climate information. SWITCH-ON is an European venture that works towards manageable utilization of water assets, a sheltered society and headway of hydrological sciences in view of Open Data. The venture expects to assemble the first stop shop gateway of open information, water data and its clients in a single place [36]. Earth Cube is a venture propelled in 2011 that builds up a typical cyber infrastructure to collect, getting to, breaking down, sharing and imagining all types of information and related assets for comprehension and anticipating a complex and developing strong Earth, hydrosphere, air, space condition frameworks, using progressed mechanical and computational abilities [37]. The on-going development of open information can help the information based research and the information utilization by evacuating the legitimate limitation on the information utilize. Numerous information entryways are being made for information sharing through web benefit with much intense information seek apparatuses where clients can discover information by area, time, and information sorts, and so on.





### **3.5. Lifts from distributed computing**

The instruments created in the huge information period, for example, Hadoop MapReduce, Apache Spark, can deal with greatly substantial datasets inside middle of the road runtime, yet the learning and system to set up and deal with the apparatuses are required. The business distributed computing, administration is accessible to researchers as an option [8], where information stockpiling and handling should be possible in the cloud, for example, Microsoft Azure, Amazon Elastic Compute Cloud, Google Compute Engine, Rackspace, Verizon and GoGrid. The business cloud has an utilization based value arrangement, influencing the figuring to work practical than executing neighborhood bunches. The distributed computing is versatile to suit the activity, and does not require broad information on designing nearby groups. US National Oceanic and Atmospheric Administration (NOAA) has propelled its Big Data Project working together with Amazon Web Service, Google Cloud Platform, IBM, Microsoft, and the Open Cloud Consortium [38]. The NOAA information will be conveyed to the cloud stage together with enormous information preparing administrations, for example, Google BigQuery and Google Cloud Dataflow, to investigate, and make new discoveries. NOAA's Big Data Project demonstrated a common pattern of brushing the colossal volume of superb information hold by the administration and the modern huge framework and specialized limit of information administration and investigation.

### **4. Conclusion**

The enormous information time is an up and coming pattern that nobody can escape from. Researchers are

relied upon to grasp the enormous information period judiciously without being obscured by the mind-boggling pattern. The idea of enormous information started from the promotion of web as digitalizing of the data among the world turns out to be considerably simpler and less expensive for a future information mining reason. The business esteem, e.g., accuracy promoting, information based basic leadership, behind the growing datasets makes the term 'enormous information' to a great degree in vogue. The possibility of huge, information is extremely versatile, and can be profitable for scholastic reason too. Hydroinformatics can profit by the consuming measure of information gathered, produced and opened to the exploration group. Information from keen meters, savvy sensors and shrewd administrations, remote detecting, earth perception frameworks, Internet of Things, and so on., will incite hydroinformatics into the inescapable enormous information time. The information utilization can be sorted into three measurements, the normal measurement, investigating the environmental change, surge and dry season administration and the worldwide water cycle; the social measurement, concentrating on the connection between water condition and the human culture; and the business measurement, utilizing information based basic leadership framework for upgrading the water asset administration framework and future water arranging. The information is prepared apparatuses like parallel processing, circulated capacity have been created to help clients to deal with the extensive data sets in hundreds GBs or TBs in passable time to make ongoing application conceivable and intuitive human-PC investigation plausible. The distributed computing stages will make it superfluous to download the information to nearby machine or run the model locally, however give prevalent registering



effectiveness later on distributed computing period. The genuine test soon is the way to make the best utilization of the accessible information, as at present there is minimal done about huge information pertinent to hydroinformatics. In this manner, the motivation behind the paper is to urge the examination group to grow new thoughts for the huge information time.

## References

- [1] Abbott, M., *Introducing Hydroinformatics. Journal of hydroinformatics*, 1999. 1: p. 3-19.
- [2] Abbott, M.B., *Hydroinformatics: information technology and the aquatic environment*. 1991: Avebury Technical.
- [3] Apache. *What Is Apache Hadoop?* 2015 [cited 2015 June 22]; Available from: <https://hadoop.apache.org/>.
- [4] Bennett, J. and S. Lanning. *The netflix prize. In Proceedings of KDD cup and workshop*. 2007. New York: ACM.
- [5] Burn-Murdoch, J. *Study: less than 1% of the world's data is analysed, over 80% is unprotected*. 2012 [cited 2015 June 18 ]; Available from: <http://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume>.
- [6] Commerce, D.o. *U.S. Secretary of Commerce Penny Pritzker Announces New Collaboration to Unleash the Power of NOAA's Data*. 2015 [cited 2015 02 Dec]; Available from: <https://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerce-penny-pritzker-announces-new-collaboration-unleash>.
- [7] Council, N.R., *Frontiers in massive data analysis*. 2013, Washington, D.C.: The National Academies Press.
- [8] Dadi, Sanyasinaidu. "USE OF GIS IN HYDROLOGICAL INVESTIGATIONS." *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED RESEARCH TRENDS*.(2015): 67. Web. <[https://www.researchgate.net/publication/321069487\\_USE\\_OF\\_GIS\\_IN\\_HYDROLOGICAL\\_INVESTIGATIONS](https://www.researchgate.net/publication/321069487_USE_OF_GIS_IN_HYDROLOGICAL_INVESTIGATIONS)>.
- [9] Dadi, Sanyasinaidu. "An Importunate Role of GIS in Indian Retail Industry." *MAT Journals-Journal of Remote Sensing GIS & Technology*. (2017): Web. <<http://www.matjournals.in/index.php/JORSGT/article/view/2110>>.
- [10] Dadi, Sanyasinaidu. "Remote Sensing and Geographic Information System for Jungle Administration." *MAT -matjournals-JOURNAL OF REMOTE SENSING GIS & TECHNOLOGY* (2017): Web. <<http://matjournals.in/index.php/JoADC/article/view/2128>>.
- [11] Dadi, Sanyasinaidu. "GIS and Remote Sensing as Tool to Develop Applications for Natural Resource Management." *MAT -matjournals-JOURNAL OF REMOTE SENSING GIS & TECHNOLOGY*. (2017): Web. <<http://www.matjournals.in/index.php/JORSGT/article/view/2101>>.
- [12] Dadi, Sanyasinaidu. "GIS and Remote Sensing For Site Specific Farming Area Mapping." *MAT-matjournals-Journal of Analog and Digital Communications*. (2017): Web. <<http://matjournals.in/index.php/JoADC/article/view/2126>>.
- [13] Dadi, Sanyasinaidu. "UNDERSTANDING THE CONCEPT OF VIRTUAL GLOBE FOR A GIS PERSONNEL." *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED RESEARCH TRENDS*. (2015): Web. <<http://ijmart.in/PreviousIssues/Sep-2015/4.pdf>>.
- [14] Dadi, Sanyasinaidu. "GIS Applications to Smart Cities." *International Journal of Advanced Multidisciplinary Scientific Research(IJAMSR) Volume 1, Issue 1, February* (2018).
- [15] Dadi Sanyasinaidu, and Peddada Jagadeeswara Rao. "Study on Sustainable Management of Groundwater Resources in Greater Visakhapatnam Municipal Corporation, Visakhapatnam District, India—A Hydro Informatics Approach." *Springer Series in Geomechanics and Geoengineering Proceedings of International Conference on Remote Sensing for Disaster Management*, 2018, pp. 719–728., doi:10.1007/978-3-319-77276-9\_64.
- [16] De Mauro, A., M. Greco, and M. Grimaldi. *What is Big Data? A Consensual Definition and a Review of Key Research Topics*. in *4th International Conference on Integrated Information, Madrid*. doi. 2014.
- [17] Dean, J. and S. Ghemawat, *MapReduce: simplified data processing on large clusters*. *Communications of the ACM*, 2008. 51(1): p. 107-113.
- [14] Ghemawat, S., H. Gobioff, and S.-T. Leung. *The Google file system*. in *ACM SIGOPS operating systems review*. 2003. ACM.
- [18] ECMWF. *ECMWF Climate Reanalysis*. 2015 [cited 2015 July 16]; Available from: <http://www.ecmwf.int/en/research/climate-reanalysis>. [27]. Kobayashi, S., Y. Ota, and Y. Harada, *The JRA-55 Reanalysis: General Specifications and Basic Characteristics*. *Journal of the Meteorological Society of Japan*, 2015. 93(1): p. 5-48.
- [19] Eilander, D. *Twitter used to create real-time flood maps. Deltares 201* [cited 2015 April 27]; Available from:



- [20]Gantz, J. and D. Reinsel, *The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east*. IDC iView: IDC Analyze the Future, 2012. 2007: p. 1-16.
- [21]Ginsberg, J., et al., *Detecting influenza epidemics using search engine query data*. *Nature*, 2009. 457(7232): p. 1012-4. Lazer, D., et al., *Big data. The parable of Google Flu: traps in big data analysis*. *Science*, 2014. 343(6176): p. 1203-5. [5].Butler, D., *When Google got flu wrong*. *Nature*, 2013. 494(7436): p. 155.
- [22]Hartin, E. and K. Watson. *Announces New Innovations that Set the Standard for Performance, Reliability, Capacity, Agility and Efficiency for Helping Companies Harness the Power of Data*. HGST Storage 2014
- [23]<https://www.deltares.nl/en/news/twitter-used-to-create-real-time-flood-maps/>.
- [24]ITU. *Internet of Things Global Standards Initiative*. 2015
- [25]Lampos, V. and N. Cristianini, *Nowcasting events from the social web with statistical learning*. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2012. 3(4): p. 72.
- [26]Lansdall-Welfare, T., et al. *On the coverage of science in the media: A big data study on the impact of the Fukushima disaster*. in *Big Data(Big Data)*, 2014 IEEE International Conference on. 2014. IEEE.
- [27]Leicester, U.o. *Big data technology finds ideal river locations to generate hydro-power*.
- [28]Mayer-Schönberger, V. and K. Cukier, *Big data: A revolution that will transform how we live, work, and think*. 2013: Houghton Mifflin Harcourt.
- [29]NASA. *Global Precipitation Measurement (GPM) Mission Overview*. 2011
- [30]National Centers for Environmental Prediction, N.W.S.N.U.S.D.o.C., NCEP/NCAR Global Reanalysis Products, 1948-continuing. 1994, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory: Boulder, CO.
- [31]Open definition. *Defining Open in Open Data, Open Content and Open Knowledge*.
- [32]Pierson, L. *Civil Engineer Turned Environmental Data Scientist Harnesses Big Environmental Data at UNESCO-IHE*. 2014
- [33]Selding, P.B.d. *U.S. Government-leased Satellite Capacity Going Unused*. 2012
- [34]Snijders, C., U. Matzat, and U.-D. Reips, *Big data: Big gaps of knowledge in the field of internet science*. *International Journal of Internet Science*, 2012. 7(1): p. 1-5.
- [35]Zaharia, M., et al. *Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing*. in *Proceedings of the 9<sup>th</sup> USENIX conference on Networked Systems Design and Implementation*. 2012. USENIX Association.