Big Data and Cloud Computing: Innovation Opportunities and Challenges

Gaurav Khandelwal, Research Scholar, Calorx Teachers' University, Ahmedabad, Gujarat.
Dr. Kunwar Singh Vaisla, Associate Professor & Head, Department of Computer Science & Engineering, BT Kumaon Institute of Technology, Dwarahat - 263653, Dist - Almora, Uttarakhand.

ABSTRACT

Large Data has developed in the previous hardly any years as another worldview giving copious information and chances to improve as well as empower research and choice help applications with exceptional incentive for digital earth applications including business, sciences and building. Simultaneously, Big Data presents difficulties for digital earth to store, transport, procedure, mine and serve the information. Distributed computing gives major help to address the difficulties with shared computing assets including computing, stockpiling, organizing and logical programming; the utilization of these assets has encouraged noteworthy Big Data headways. This paper studies the two boondocks – Big Data and distributed computing – and audits the favorable circumstances and results of using distributed computing to handling Big Data in the digital earth and important science areas. From the parts of a general presentation, sources, challenges, innovation status and research openings, the accompanying perceptions are offered: (I) distributed computing and Big Data empower science disclosures and application improvements; (ii) distributed computing gives significant answers for Big Data; (iii) Big Data, spatiotemporal reasoning and different application spaces drive the progression of distributed computing and pertinent advancements with new prerequisites; (iv) inborn spatiotemporal standards of Big Data and geospatial sciences give the source to discovering specialized and hypothetical answers for advance distributed computing and preparing Big Data; (v) open accessibility of Big Data and handling capacity present social difficulties of geospatial importance and (vi) a mesh of developments is changing Big Data into geospatial research, building and business esteems. This audit presents future developments and an examination plan for distributed computing supporting the change of the volume, speed, assortment and veracity into estimations of Big Data for neighborhood to worldwide digital earth science and applications.

KEYWORDS: Spatiotemporal computing, digital earth, geospatial cyberinfrastructure, geoinformatics, CyberGIS, Smart cities

1. Introduction

Big Data refers to the flood of digital data from many digital earth sources, including sensors, digitizers, scanners, numerical modeling, mobile phones, Internet, videos, e-mails and social networks. The data types include texts, geometries, images, videos, sounds and combinations of each. Such data can be directly or indirectly related to geospatial information (Berkovich and Liao 2012). The evolution of technologies and human understanding of the data have shifted data handling from the more traditional static mode to an accelerating data arena characterized by volume, velocity, variety, veracity and value (i.e. 5Vs of Big Data; Marr 2015). The first V refers to the volume of data which is growing explosively and extends beyond our capability of handling large data sets; volume is the most common descriptor of Big Data (e.g. Hsu, Slagter, and Chung 2015). Velocity refers to the fast generation and transmission of data across the Internet as exemplified by data collection from social networks, massive array of sensors from the micro (atomic) to the macro (global) level and data transmission from sensors to supercomputers and decision-makers. Variety refers to the diverse data forms and in which model and structural data are archived. Veracity refers to the diversity of quality, accuracy and trustworthiness of the data. All four Vs are important for reaching the 5th V, which focuses on specific research and decision-support applications that improve our lives, work and prosperity (Mayer-Schönberger and Cukier 2013).

2. Sources of Big Data

While Big Data has the component of 5Vs, the element based difficulties shift in various digital earth pertinent spaces. This segment audits important space explicit Big Data challenges in the succession of how intently are they identified with geospatial standards .

2.1. Earth sciences

The progression of detecting and computing recreation advancements empowered assortment and age of huge informational collections consistently at various spatiotemporal scales for checking, understanding and introducing complex earth frameworks. For instance, EO gathers TB of pictures day by day (Yang et al. 2011a) with expanding spatial (for example sub-meter), worldly (hourly) and ghostly (many groups) goals (Benediktsson, Chanussot, and Moon 2013). Geospatial models likewise create enormous spatiotemporal information by means of numerical reenactments of the intricate earth frameworks. Atmosphere science is a model speaking to the Big Data move over all digital earth areas (Edwards 2010; Schnase et al. 2014) in utilizing huge spatiotemporal information to screen and depict the perplexing earth atmosphere framework. For instance, the IPCC AR5 alone delivered 10,000 TB of atmosphere information, and the following IPCC will draw in several PB. It is basic to proficiently dissect these information for recognizing worldwide temperature oddities, distinguishing geological areas with comparative or unique atmosphere examples, and examining spatiotemporal dissemination of extraordinary climate occasions (Das and Parthasarathy 2009). Be that as it may, productively mining data from PB of atmosphere information is as yet testing (Li 2015).

2.2. Internet of Things

Propelled sensors and their facilitating gadgets (for example cell phones, wellbeing screens) are associated in a digital physical framework to gauge time and area of people, development of vehicles, vibration of machine, temperature, precipitation, moistness and compound changes in the air (Lohr 2012). The Internet of Things (IoT, Camarinha-Matos, Tomic, and Graça 2013) catches this new area and consistently produces information streams over the globe with land impressions from interconnected cell phones, PCs, sensors, RFID labels and cameras (Michael and Miller 2013; Van sanctum Dam 2013).

2.3. Social sciences

Informal organizations, for example, Twitter and Facebook, produce Big Data and are changing sociologies. As of the hour of composing, Twitter clients around the globe are delivering around 6000 tweets for every subsequent which compares to 500 million tweets for each day and around 200 billion tweets for each year (Internet Live Stats 2016). Financial specialists, political researchers, sociologists and other social researchers utilize Big Data mining techniques to break down social collaborations, wellbeing records, telephone logs, government records and other digital follows (Boyd and Crawford 2012).

2.4. Astronomy

Space science is delivering a spatiotemporal guide of the universe (Dillon 2015) by watching the sky utilizing propelled sky study advances. Mapping and looking over the universe produces immense measures of spatiotemporal information. For instance, Sloan Digital Sky Survey (SDSS) produced 116 TB information for its Data Release 12 (Alam et al. 2015). New observational instruments (for example Huge Synoptic Survey Telescope) booked for activity in 2023 will create 15 TB information daily and convey 200 PB information altogether to address the structure and development of the universe (http://www.lsst.org).

2.5. Business

Business knowledge and examination are upgraded with Big Data for choices on technique, overseeing enhancement and rivalry (Chen, Chiang, and Story 2012; Gopalkrishnan et al. 2012). Business activities (for example Visa changes, online buys) create enormous volume, high speed and exceptionally unstructured (assortment and veracity) informational indexes. These information contain rich geospatial data, for example, where and when a progress happened. To oversee and process these information, the full range of information preparing advances has been created for the appropriated and versatile stockpiling condition (Färber et al. 2012; Indeck and Indeck 2012; Moniruzzaman and Hossain 2013).

3. Big Data technology challenges

While following the existence cycle difficulties of conventional information, digital earth Big Data presents other innovative difficulties as a result of its 5V includes in a wide range of segments of industry, government and technical studies (McAfee et al. 2012). This segment surveys the mechanical difficulties presented by Big Data.

3.1. Data storage

Capacity challenges are presented by the volume, speed and assortment of Big Data. Putting away Big Data on conventional physical stockpiling is dangerous as hard circle drives (HDDs) frequently come up short, and customary information insurance instruments (for example Strike or repetitive cluster of autonomous circles) are not effective with PB-scale stockpiling (Robinson 2012). What's more, the speed of Big Data requires the capacity frameworks to have the option to scale up rapidly which is hard to accomplish with conventional stockpiling frameworks. Distributed storage administrations (for example Amazon S3, Elastic Block Store or EBS) offer basically boundless stockpiling with high adaptation to non-critical failure which gives potential answers for address Big Data stockpiling difficulties.

3.2. Data transmission

Information transmission continues in various phases of information life cycle as follows: (I) information assortment from sensors to capacity; (ii) information reconciliation from different server farms; (iii) information the board for moving the incorporated information to handling stages (for example cloud stages) and (iv) information investigation for moving information from capacity to breaking down host (for example superior computing (HPC) groups). Moving enormous volumes of information presents evident difficulties in every one of these stages.

3.3. Data management

It is hard for PCs to effectively oversee, dissect and envision enormous, unstructured and heterogeneous information. The assortment and veracity of Big Data are reclassifying the information the board worldview, requesting new innovations (for example Hadoop, NoSQL) to clean, store, and sort out unstructured information (Kim, Trimi, and Chung 2014). While metadata are basic for the honesty of information provenances (Singh et al. 2003; Yee et al. 2003), the test stays to naturally produce metadata to depict Big Data and pertinent procedures (Gantz and Reinsel 2012; Oguntimilehin and Ademola 2014).

3.4. Data processing

Preparing huge volumes of information requires devoted computing assets and this is incompletely taken care of by the speeding up CPU, system and capacity (Bertino et al. 2011). Anyway the computing assets required for handling Big Data far surpass the preparing power offered by conventional driving standards (Ammn and Irfanuddin 2013). Distributed computing offers practically boundless and on-request preparing power as an incomplete arrangement. In any case, moving to the cloud introduces various new issues. First is the constraint of distributed computing's system transfer speed which impacts the calculation productivity over enormous information volumes (Bryant, Katz, and Lazowska 2008).

3.5. Data analysis

Information examination is a significant stage in the worth chain of Big Data for data extraction and expectations (Fan and Liu 2013; Chen et al. 2014b). In any case, investigating Big Data challenges the intricacy and adaptability of the hidden calculations (Khan et al. 2014). Huge Data investigation requires modern adaptable and interoperable calculations (Jagadish et al. 2014) and is tended to by welding examination projects to parallel handling stages (for example Hadoop) to bridle the intensity of circulated handling. Be that as it may, this 'separate and vanquish' procedure doesn't work with profound and multi-scale cycles (Chen and Zhang 2014) that are required for most geospatial information examination/mining calculations.

3.6. Data visualization

Large Data representation reveals concealed examples and finds obscure connections to improve basic leadership (Nasser and Tariq 2015). Since Big Data is frequently heterogeneous in type, structure and semantics, representation is basic to understand Big Data (Chen et al. 2014b; Padgavankar and Gupta 2014). However, it is hard to give ongoing representation and human cooperation for outwardly investigating and dissecting Big Data (Sun et al. 2012; Jagadish et al. 2014; Nasser and Tariq 2015).

3.7. Data integration

Information reconciliation is basic for accomplishing the fifth V (estimation) of Big Data through integrative information investigation and cross-space joint efforts (Chen et al. 2013; Christen 2014). Dong and Divesh (2015) outlined the information coordination difficulties of pattern mapping, record linkage and information combination. Metadata is basic for following these mappings to make the incorporated information sources 'mechanically' resolvable and to encourage enormous scale examinations (Agrawal et al. 2011). Anyway effectively and naturally making metadata from Big Data is as yet a difficult assignment (Gantz and Reinsel 2011).

3.8. Information engineering

Large Data is step by step changing the manner in which logical research is led as confirm by the inexorably information driven and the open science approach (Jagadish et al. 2014). Such changes present difficulties to framework engineering. For instance, flawlessly incorporating various devices and geospatial administrations stay a high need (Li et al. 2011; Wu et al. 2011). Extra need issues incorporate coordinating these devices into reusable work processes (Li et al. 2015b), fusing information with the devices to advance usefulness (Li et al. 2014) and sharing information and investigations among networks.

3.9. Information security

The expanding reliance on PCs and Internet over the previous decades makes organizations and people defenseless against information rupture and misuse (Denning and Denning 1979;

Abraham and Paprzycki 2004; Redlich and Nemzow 2006). Huge Data presents new security challenges for customary information encryption norms, procedures and calculations (Smid and Branstad 1988; Coppersmith 1994; Nadeem and Javed 2005).

3.10. Information security challenges

The extraordinary systems administration among smart gadgets and computing stages adds to Big Data however presents protection concerns where a person's area, conduct and exchanges are digitally recorded (Cukier 2010; Tene 2011; Michael and Miller 2013; Cheatham 2015). For instance, online life and individual medicinal records contain individual wellbeing data raising protection concerns (Terry 2012; Kaisler et al. 2013; Michael and Miller 2013; Padgavankar and Gupta 2014).

3.11. Information quality

Information quality incorporates four viewpoints: precision, fulfillment, excess and consistency (Chen et al. 2014b). The inherent idea of multifaceted nature and heterogeneity of Big Data makes information precision and fulfillment hard to distinguish and follow, along these lines expanding the danger of 'bogus revelations' (Lohr 2012). For instance, web based life information are profoundly slanted in space, time and socioeconomics, and area exactness differs from meters to many kilometers.

4. Distributed computing and other applicable innovation scene

This area surveys the innovation challenges presented by Big Data from 12 distinct perspectives. While a portion of these difficulties, (for example, investigation, perception and quality) exist before Big Data period, the 5Vs of Big Data carry the difficulties to another level as talked about above. Large Data presents interesting difficulties from a few angles including examination, perception, joining and design, because of the natural high-dimensionality of geospatial information and the complex spatiotemporal connections.

4.1. Information stockpiling, the executives and model

4.1.1. Disseminated record/stockpiling framework

To address the capacity difficulty, an expanding number of disseminated record frameworks (DFSs) are adjusted with capacity of little documents, load adjusting, duplicate consistency and de-duplication (Zhang and Xu 2013) in a system shared records and capacity design (Yeager 2003). The Hadoop Distributed File System (HDFS; Shvachko et al. 2010) is such a framework running on various hosts, and numerous IT organizations, including Yahoo, Intel and IBM, have embraced HDFS as the Big Data stockpiling innovation.

4.1.2. NoSQL database framework

While contemporary information the board arrangements offer constrained combination abilities for the assortment and veracity of Big Data, ongoing advances in distributed computing and NoSQL open the entryway for new arrangements (Grolinger et al. 2013). The NoSQL databases coordinate necessities of Big Data with high adaptability, accessibility and adaptation to non-critical failure (Chen et al. 2014a). Numerous examinations have researched developing Big Data innovations (for example MapReduce structures, NoSQL databases) (Burtica et al. 2012). A Knowledge as a Service (KaaS) structure is proposed for debacle cloud information the board, where information are put away in a cloud situation utilizing a blend of social and NoSQL databases (Grolinger et al. 2013).

4.1.3. Search, inquiry, ordering and information model plan

Execution is basic in Big Data time, and precisely and rapidly finding information requires another age of web crawlers and question frameworks (Miyano and Uehara 2012; Aji et al. 2013). Zhong et al. (2012) proposed a way to deal with give effective spatial inquiry preparing over huge spatial information and various simultaneous client inquiries. This methodology first arranges spatial information in quite a while of geographic nearness to accomplish high Input/Output (I/O) throughput, at that point plans a two-level disseminated spatial file for pruning the hunt space, lastly utilizes an 'indexing + MapReduce' information preparing design for productive spatial question.

4.2. Information preparing, mining and information disclosure

4.2.1. MapReduce (Hadoop) framework

MapReduce is a parallel programming model for Big Data with high versatility and adaptation to non-critical failure. The rich structure of MapReduce has incited the usage of MapReduce in various computing designs, including multi-center bunches, mists, Cubieboards and GPUs (Jiang et al. 2015). MapReduce likewise has gotten an essential decision for cloud suppliers to convey information systematic administrations (Zhao et al. 2014). Numerous conventional calculations and information preparing in a solitary machine condition are moved to the MapReduce stage (Kim 2014; Cosulschi, Cuzzocrea, and De Virgilio 2013). For instance, Kim (2014) broke down a calculation to produce wavelet summaries on the appropriated MapReduce system.

4.2.2. Parallel programming dialects

While parallel computing (for example MapReduce) is generally utilized and viable in Big Data, a dire need is successful programming models and dialects (Hellerstein 2010; Dobre and Xhafa 2014). Alvaro et al. (2010) exhibited that explanatory dialects significantly streamline conveyed frameworks programming. Datalog-motivated dialects have been investigated with an attention on naturally parallel errands in systems administration and circulated frameworks and have demonstrated to be an a lot less complex group of dialects for programming parallel and dispersed programming (Hellerstein 2010).

4.2.3. Factual examinations, AI and information mining

Standard factual and information digging apparatuses for conventional informational indexes are not intended for supporting measurable and AI examination for Big Data (Triguero et al. 2015) in light of the fact that numerous conventional apparatuses (for example R) just sudden spike in demand for a solitary PC. Numerous researchers have examined parallel and versatile computing to help the most regularly utilized calculations.

4.2.4. Huge Data investigation and representation

Huge Data investigation is a rising examination subject with the accessibility of monstrous stockpiling and computing abilities offered by cutting edge and versatile computing frameworks. Baumann et al. (2016) presented the EarthServer, a Big Earth Data Analytics motor, for inclusion type datasets dependent on superior exhibit database innovation, and interoperable principles for administration communication (for example OGC WCS and WCPS).

4.2.5. Semantics and metaphysics driven methodologies

Semantic and ontologies make PC and web smarter to comprehend, control and break down an assortment of information. Semantic has risen as a typical, reasonable information model that encourages interoperability, joining and checking of information based frameworks (Kourtesis, Alvarez-Rodríguez, and Paraskakis 2014). Lately there has been a blast of enthusiasm utilizing semantics for the conventional information investigation since it gets the genuine data from information. Particularly for cross-area information, semantic gives points of interest to connect information and trade data.

4.3. Portable information assortment, computing and Near Field Communication

Cell phones are assuming a noteworthy job in our lives, and cell phones enable applications to gather and use Big Data and escalated computing assets (Soyata et al. 2012). Truth be told, enormous volumes of information on gigantic scales have been created from GPS-prepared cell phones, and related strategies and approaches are generally created to join such information for different applications, for instance, human portability contemplates, course.

4.4. Enormous Data computing and handling foundation

The highlights of Big Data drive research to find far reaching answers for both computing and information handling, for example, structuring propelled models, and creating information segment and parallelization techniques to all the more likely influence HPC. Liu (2013) reviewed computing framework for Big Data handling from parts of design, stockpiling and systems administration challenges and talked about developing computing foundation and innovations.

4.5. Large Data and cloud answers for geospatial sciences

While distributed computing developed as a potential answer for help computing and information concentrated applications, a few obstructions prevent progressing from customary computing to distributed computing (Li et al. 2015b). In the first place, the expectation to absorb information for geospatial researchers is steep with regards to understanding the administrations, models and cloud procedures. Second, inherent difficulties brought by cloud foundation (for example correspondence overhead, ideal cloud zones) presently can't seem to be tended to. What's more, third, handling gigantic information and running models includes complex techniques and different instruments, requiring progressively adaptable and helpful cloud administrations. This segment examines the arrangements used to all the more likely influence distributed computing for tending to Big Data issues in the geospatial sciences.

4.5.1. Anything as a service (XaaS) for geospatial data and models

There are rising highlights to facilitate the test of utilizing distributed computing and encourage the utilization of generally accessible models and administrations. Eminent models are web administration structure (Liu et al. 2015), web demonstrating (Geller and Turner 2007), application as a help (Lushbough, Gnimpieba, and Dooley 2015), MaaS (Li et al. 2014) and work process as an assistance (WaaS; Krämer and Senner 2015).

4.5.2. Computing asset auto-arrangement, scaling and planning

The huge scale information prerequisite of joint effort, productive information preparing and the regularly expanding divergent types of client applications are making information the board progressively intricate as well as are bringing more difficulties for the asset auto-arrangement, scaling and planning of the underneath computing framework (Gui et al. 2016). With the improvement of cloud advancements and broad organization of cloud stages, computing offices are for the most part made in the configuration of virtualized assets or virtual machines (VMs).

4.5.3. Enhancing Big Data stages with spatiotemporal standards

The advancement procedure for planning, creating and sending geospatial applications utilizing spatiotemporal standards are examined in Yang et al. (2011a, 2011b). Spatiotemporal examples can be seen from the accompanying: (I) physical area of computing and capacity assets; (ii) dispersion of information; (iii) dynamic and gigantic simultaneous access of clients at various areas and times and (iv) study region of the applications.

5. Current status of handling Big Data challenges with distributed computing

While the Big Data difficulties can be handled by many trend setting innovations, for example, HPC, distributed computing is the most tricky and significant. This segment surveys the status of utilizing distributed computing to address the Big Data challenges.

5.1. On-request asset arrangement

The volume and speed difficulties of Big Data require VM creation on-request. Self-sufficient recognition of the speed for provisioning VMs is basic (Baughman et al. 2015) and ought to think about both ideal expense and high effectiveness in task execution (Pumma, Achalakul, and Li 2012).

5.2. Booking

Occupation booking successfully allots computing assets to a lot of various assignments. Nonetheless, booking is a test in programmed and dynamic asset provisioning for Big Data (Vasile et al. 2014; Gui et al. 2016). Zhan et al. (2015) proposed a few research bearings for cloud asset planning, including continuous, versatile dynamic, huge scale, multi-objective, and dispersed and parallel booking. Sfrent and Pop (2015) indicated that under specific conditions one can find the best booking calculation. Hung, Aazam, and Huh (2015) introduced an expense and time-mindful hereditary booking calculation to streamline execution.

5.3. Adaptability

Adaptability on circulated and virtualized processors is and has been a bottleneck for utilizing distributed computing to process Big Data. Feller, Ramakrishnan, and Morin (2015) explored cloud versatility and closed: (I) existing together VMs decline the circle throughput; (ii) execution on physical bunches is altogether superior to that on virtual groups; (iii) execution debasement because of division of the administrations relies upon the information to-process proportion and (iv) application fruition progress corresponds with the force utilization, and force utilization is application explicit. Different cloud execution benchmarks and assessments (Nazir et al. 2012; Huang et al. 2013b) exhibited that adjusting the number and size of VMs as an element of the particular applications is basic to accomplish ideal adaptability for geospatial Big Data applications. Ku, Choi, and Min (2014) dissected four cloud execution affecting components (for example number of inquiry proclamations, trash assortment interims, amount of VM assets and virtual CPU task types).

5.4. Information area

Framework I/O represents a bottleneck for Big Data handling in the cloud (Kim et al. 2014) particularly when information and computing are topographically scattered. Analysts either move information to computing asset or move the computing assets to the information (Jayalath, Stephen, and Eugster 2014).

5.5. Distributed computing for web based life and other spilled information

The assortment and veracity of web based life and other spilled information present new difficulties to the contemporary information handling and capacity structures and models. For

Big Data the executives, numerous non-customary philosophies, for example, NoSQL and versatile SQL are actualized (Nambiar, Chitor, and Joshi 2014).

6. Research plan

Huge Data systems and significant difficulties must be supported by a substantive research activity. A few examinations have recognized new activities for Big Data and distributed computing (Bughin, Chui, and Manyika 2010; Karimi 2014; Assunção et al. 2015; Hashem et al. 2015) from various features of the executives/design, model advancement, perception, association and business. To completely use and propel distributed computing for Big Data control, inquire about is required to advance from the volume, speed, assortment and veracity of Big Data to the estimation of the items, and these examination activities are portrayed underneath.

6.1. Dispersed information stockpiling and the executives challenges

6.2. Huge Data mining

6.3. Hazard the board and quality affirmation

6.4. Spatiotemporal collocation

6.5. Interdisciplinary coordinated effort

6.6. Work process partaking in business, science and application rationale support

6.7. Half breed approach and anything as a help

6.8. Determination, actuation and valuing

6.9. Give reasonable and accessible training

References

- Abbas, A., K. Bilal, L. Zhang, and S. U. Khan. 2015. "A Cloud Based Health Insurance Plan Recommendation System: A User Centered Approach." Future Generation Computer Systems 43–44: 99–109. [Crossref], [Web of Science ®], [Google Scholar]

- Abolfazli, S., Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya. 2014. "Cloud-Based Augmentation for Mobile Devices: Motivation, Taxonomies, and Open Challenges." IEEE Communications Surveys Tutorials 16 (1): 337–368. [Crossref], [Web of Science ®], [Google Scholar]

- Abouzeid, A., K. Bajda-Pawlikowski, D. Abadi, A. Silberschatz, and A. Rasin. 2009. "HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for

Analytical Workloads." Proceedings of the VLDB Endowment 2 (1): 922–933. [Crossref], [Google Scholar]

- Abraham, An., and M. Paprzycki. 2004. "Criticalness of Steganography on Data Security." In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04). Vol. 2, 347–351. IEEE. [Google Scholar]

- Aghabozorgi, S., A. Seyed Shirkhorshidi, and T. Ying Wah. 2015. "Time-arrangement Clustering – A Decade Review." Information Systems 53 (C): 16–38. [Crossref], [Web of Science ®], [Google Scholar]

- Agrawal, D., P. Bernstein, E. Bertino, S. Davidson, U. Dayal, M. Franklin, J. Gehrke, et al. 2011. Difficulties and Opportunities with Big Data 2011–1. Digital Center Technical Reports. http://docs.lib.purdue.edu/cctech/1. [Google Scholar]

- Agrawal, D., S. Das, and A. El Abbadi. 2011. "Enormous Data and Cloud Computing: Current State and Future Opportunities." In Proceedings of the fourteenth International Conference on Extending Database Technology, 530–533. ACM. [Google Scholar]

- Aji, A., F. Wang, H. Vo, R. Lee, Q. Liu, X. Zhang, and J. Saltz. 2013. "Hadoop GIS: A High Performance Spatial Data Warehousing System Over Mapreduce." Proceedings of the VLDB Endowment 6: 1009–1020. [Crossref], [Google Scholar]
- Alam, S., F. D. Albareti, C. A. Prieto, F. Anders, S. F. Anderson, B. H. Andrews, E. Armengaud, et al. 2015. "The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III." The Astrophysical Journal Supplement Series 219 (1): 1–27. [Crossref], [Web of Science ®], [Google Scholar]
- Alvaro, P., T. Condie, N. Conway, K. Elmeleegy, J. M. Hellerstein, and R. Burns. 2010. "Blast Analytics: Exploring Data-driven, Declarative Programming for the Cloud." In Proceedings of the fifth European Conference on Computer Systems, 223–236. New York, NY: ACM. [Google Scholar]