Domains and Methodologies for Big Data Project in Software Engineering

Dr. Sanjeev Punia[1], Mr. Manoj Kumar[2], Dr. Kuldeep Malik[3]

[1]Department of Computer Sceince, JIMSEMTC, G.Noida (India), [2]Department of Computer Science, The Northcap University, Gurugram (India), [3]Department of Computer Sceince, ITS Engineering College, G.Noida (India),

[1]drsanjeevpunia@hotmail.com, [2]wss.manojkumar@gmail.com, [3]kuldeepmalik@its.edu.in

Abstract- Nowadays big data became the new buzzword in the field of information and communication technology. Presently, researchers are looking to extract the maximum value of big data applications from available big data. However, developing, maintaining and scaling the big data application is still a distant milestone. We build better big data application projects by using existing software development life cycle (SDLC) phase in software engineering. The result of research helped in identifying big data application potential projects that utilize big data successfully. The proposed paper helps in exploring software development life cycle (SDLC) phases in big data applications and perform phase gap analysis to find the detailed efforts in research.

Keywords**:** *Big data patterns, Application projects, Data methodologies.*

1 INTRODUCTION

There are several success stories of big data that is used by technology giants to dominate their competitors in areas like social media, search engines, e-commerce and video streaming services etc. Some popular players and leaders in the field of big data are Facebook, LinkedIn, Twitter, Google, Amazon and Netflix. The success of these big data user piqued the interest of numerous companies to extract maximum value from the available data.

Gartner et.al. explained that 64% survey respondents planned to invest in big data applications in 2013 where only less than 8% had actually deployed at the time of the survey. The case study of digital manufacturing process optimization shows the possibility to enhance process efficiency using big data. Big data is used in different sectors such as healthcare, trading, agriculture, tourism and politics for stakeholder's advantage. Enriching customer experience using big data has been exemplified by the "people you may know" option offered by Facebook and LinkedIn and even the "customers who bought this item should also bought" service provided by Amazon.

Big data has been characterized by 4V's as Volume, Velocity, Variety and Veracity. Volume implies the data explosion mark of computing world in the last decade. Velocity is the constraint that demands real time processing of available data where failure results in its loss or obsolescence. Variety means data format could be structured, semi-structured or unstructured and have multiple source applications those must be capable of handling these sources. Finally, Veracity implies that the cleaning of available data, historical data and real-time streaming data prior to processing to ensure its application usefulness. The whole process of developing software is fraught with errors which arise due to communication requirement changes. This is true for ordinary software development applications despite the fact that we have been developing software for more than 20 years. The complexity increases more in developing big data application software by taking the 4V's factors into account.

The software development life cycle (SDLC) concepts must be leveraged to build robust and scalable big data application projects despite the involved complexities. The building of fault tolerant big data application use the best practices and methodologies that is advantageous for both stakeholders and developers and capable of handling even more data than envisioned at the time of its creation. There have been numerous reviews on big data as the art in architecture and large scale data analysis platforms, comprehensive big data survey and related technologies acquisition in big data applications. Pakkata et.al. discussed the software engineering challenges in building data-intensive or big data software systems. However, no comprehensive study to review existing software

International Journal of Scientific & Engineering Research, Volume 8, Issue 11, ÜÎÜÚÛ-2017
ISSN 2229-5518

engineering research methodologies for enabling big data application project development.

The main goal of this paper is to look into the existing research methodologies on big data application projects. The results from this literature survey help to identify the software development life cycle (SDLC) phase domains that utilized commonly for developing the big data applications. Additional results pinpoint the domains that deploy big data application projects with advantageous outcomes but remain underexplored and definitely deserve more attention from researchers.

2 Research Method

The popular academic search engines like Scopus, Web of Science and IEEE Xplore Digital Library targeted to conduct the literature review. The search was performed using the command option under advanced search section of these search engines. A combination of keywords related to software development life cycle (SDLC) phases were selected from the software engineering like architecture, evolution, process, quality, reuse, specification, requirement, design, domain model, testing, verification, validation, maintenance, quality, analysis, framework, process and patterns.
The idea behind formulating the queries is to search the papers with software engineering and big data combined topics. We have adopted an extensive search and selection process to identify a set of possible complete studies. Our search process involved automatic and manual searching like big data requirement, specification, design, architecture, analysis, design, testing, verification/ validation, maintenance, framework, evolution and reuse domain modeling in engineering. The abstract, introduction and conclusion of each search result examined in detail is to ensure that the papers are eligible to include in the study. Admittedly, the quality of our results depends on the search queries and the efficiency of the search engines. Numerous publications from relevant big data conferences did not show in the search results because of the software development life cycle (SDLC) queries specific search terms. For example, the use of big data in the field of advertising was not covered because there is no specific software development life cycle (SDLC) phase referenced or used in this study despite the fact that the paper dealt with an interesting application domain as advertising.

A. Research Questions

The main research questions addressed throughout the study are:

RQ1. Which application domains have received attention for the development of big data application projects and which domains require more attention?

RQ2. Which software development life cycle (SDLC) phases used to enable big data applications and which fields need more research efforts?

B. Classification Criteria

The main criteria taken into account in analyzing and categorizing each paper is based on (i) paper application domain (ii) software development life cycle (SDLC) phase.

3 Limitations

This research paper reviews the results of targeted queries in popular search engines like Scopus, Web of Science and IEEE Xplore Digital Library through manual search. The main method for reviewing and selecting/discarding the papers is performed through a manual process and reproducibility of this study was not taken into consideration.

4 Results

The main purpose of this research paper is to look into the existing research of big data in software engineering and to perform a gap analysis of the research till date. The gap analysis helped to identify the application domains. The software development life cycle (SDLC) phases have not yet received much attention from researchers in big data context but have huge potential. The different application domains identified through this research are illustrated in Table I and the classification of big data software development life cycle (SDLC) phases in each paper are addressed in Table II.

A. Study Analysis

RQ1. Which application domains have received attention for the development of big data application projects and which domains require more attention?

All the application domains found by analyzing the research papers are listed in Table I. The proposed papers customized versions of existing technology without a specific domain like healthcare, military, infrastructure and information technology category. In this review, 98 papers deal with information technology domain out of 170 selected papers as shown in the count column of Table I. Nearly 57% of information technology domain analyzed papers indicates that most of the papers focus on topics that affect the computing world directly. Application domains such as healthcare, banking and financial industry are commonly believed on rich data. The healthcare industry is a source of huge data amounts sourced from the patient's electronic medical records. Data from hospitals, clinics, medical governing bodies and even insurance providers can be mined to study disease affliction

International Journal of Scientific & Engineering Research, Volume 8, Issue ŵ ŬÎŬŰŰ-2017
ISSN 2229-5518

rates, patterns and susceptibility trends. The Analyzing of medical data may help in developing innovative treatment methods, customize more effective and economical treatment plans as well as help healthcare professionals in dispensing medication for patients groups suffering from similar afflictions with identical symptoms and reactions medical history.

The banking and financial industry domain access the world monetary blueprint. The tremendous daily basis transactional amount data can be utilized by commercial banks to understand the better customers spending patterns. The use of credit card, mobile banking application, mortgage and customers credit history give great pattern of bank customers needs and help to tailor products accordingly by enhancing the customer experience and satisfaction. The metropolitan authorities (responsible for traffic management and building infrastructure facilities) can use current as well as historical data to build smart cities using minimum energy and water for maintenance and heating.

The aviation industry and the military domain may use the potential of huge big data for better aircrafts/vehicles performance and maintenance. Interestingly, geographic information systems (GIS) domain attracted the attention of big data researchers. This domain can enhance and take more benefit from the available location services data that used in building several mobile and internet applications used today.

The global warming is also the serious cause to damage our environment so wildlife and meteorologists can use global weather sensors big data to make more accurate weather predictions and provide timely natural disaster alerts.

Application Domain	Papers	Count
Information Technology	[1], [3–99]	98
Geospatial Data Processing/Geographic Information Systems	[4], [101], [113–122]	12
Environmental Monitoring/Conservation	[4], [85], [120], [149–151]	6
Healthcare	[100–112]	13
Infrastructure	[123–133]	11
Transport	[76], [123], [128], [131], 134–139]	10
Retail/Tourism/Commerce	[36], [76], [101], [140–144]	8
Social Networks	[7], [115], [116], [145–148]	7
Manufacturing	[152–154]	3
Meteorology	[118], [155], [156]	3
Cyber Physical Systems	[157–159]	3
Law & Order/Criminal Investigation/Forensic Analysis	[23], [160], [161]	3
Agriculture	[162], [163]	2
Banking and Financial Industry	[164], [165]	2
Military	[161], [166]	2
Aviation Industry	[159]	1
Astronomy	[167]	1
National Security	[161]	1

Table 1: Categories & Classification of Application Domain

RQ2. Which software development life cycle (SDLC) phases used to enable big data applications and which fields need more research efforts?

The breakdown of each software development life cycle (SDLC) phases is shown in Table II. By analyzing many papers related with software architectures, frameworks and design methodologies shows that building of big data applications primarily deal with the latest cutting edge technological developments.

Since now, there is no more research on big data software development life cycle (SDLC) phase where only some research papers explicitly explained the research through verification/validation as shown in Table II. Verification and

International Journal of Scientific & Engineering Research, Volume 8, Issue 11, ÜÎÜÚÙ2017
ISSN 2229-5518

validation processes ensure the software system requirements and design foundations written before development. Similarly, only some research papers directly deal with the maintenance phase. Some domains are extremely important to ensure the building of big data application of big data applications to meet and satisfy the goal of stakeholder without compromising performance. The big data application remains scalable with more leverage data in the future. The gathered big data need to analyze for complex process due to the 4V properties. The research requirements can provide pointers to new big data technology adopters to establish benchmarks and standard practices.

B. Classification Procedure

The research paper is classified on the basis of criteria mentioned earlier. Each paper is analyzed and categorized on the basis of application domain and software development life cycle (SDLC) phase. For example, if a paper discussed the application or implementation of big data in the healthcare industry then its application domain is classified as "Healthcare". If a paper proposed a domain specific language or described an ontology technique then its application domain is classified as "Domain Specific Language or Ontology".

Software Engineering Subfield	Papers	Count
Requirements	[17],[18],[21],[24],[26],[29],[55],[70],[74],[81],[98], [136],[142],[150],[153],[159]	16
Design	[1],[5],[13],[20],[22],[26],[30],[38],[42],[44],[47-49],[55],[57],[60],[65–67],[72],[75], [77],[84],[85],[96],[110],[123],[131],[144],[157],[159]	31
Framework	[7],[8],[11] [14-16],[25],[34],[39],[41],[45],[53],[58],[59],[62],[63],[69–71], [76],[80],[83],[88],[90],[91],[94],[100],[101],[103],[107],[109],[113],[114],[116],[122], [124–126],[132],[134],[139],[143],[145],[152],[155],[157],[158], [162],[164],[167],[173]	51
Architecture	[1],[3–6],[9],[12],[17],[22],[23],[27],[29–33],[35],[40],[43],[50],[52],[54],[56],[61],[64], [68],[73],[78],[79],[82],[85],[89],[92],[93],[95],[97],[99], [101],[102],[104–106],[110–112],[117],[119],[120],[127–129],[133],[137],[138],[141],[146–149],[151],[156],[160], [161],[162],[163],[165],[166],[174]	68
Testing	[28], [39], [43], [86], [51], [55], [79], [81], [109], [173]	10
Validation/Verification	[46], [109]	2
Maintenance	[55], [166]	2
Quality Assurance	[14], [28], [87], [98], [109], [147]	6
Domain Specific Languages/Ontology	[4], [19], [41], [53], [72], [94], [108], [121], [127], [141], [146], [149], [160]	13

Table 2: Software Development Life Cycle (SDLC) Phase

This research papers did not mention any application domain that explicitly classified under "Information Technology". If a paper proposed a framework-oriented method then it is categorized as "Framework" domain instead of "Architecture" domain.

5 Discussion

The review results of Tables I and II illustrate that there are noticeable differences in research attention amount received by different application domains and software development life cycle (SDLC) phases. The Information Technology application domain received much more

International Journal of Scientific & Engineering Research, Volume 8, Issue Ⅲ, ÜÎÜÚÛ2017
ISSN 2229-5518

research attention compared to other important and promising data rich domains such as Healthcare, Banking and Financial Industry.

There is a lot of potential to make the technology advance using identified big data domains in Aviation, Infrastructure, Transport and Environmental Monitoring/Conservation. We need to prioritize the domains that focus more on big data applications with software management and development transformation ability. The research on big data applications software methodologies help to reduce the chances of system errors, project failures and stakeholder expectations. The technologies and environment around big data evolves continuously and today big data applications deal with these unknown but inevitable changes. Similarly, more research should be conduct to enhance the existing methods, develop novel maintenance methodologies, testing, validation /verification and quality assurance of big data applications. The high risks like system unpredictability and project failure in big data applications can be mitigated by system testing, validation/verification and laying the foundations for good software quality assurance practices.

6 Conclusion and Future work

This is most likely the first comprehensive study of existing software engineering research in context of big data applications. The purpose of this research paper was to understand the big data application projects in software engineering research and highlight the use of software development life cycle (SDLC) phases in building big data applications robust and scalable.

We performed a gap analysis that identifies more popular application domains among researchers in this field. The analysis also revealed the main software development life cycle (SDLC) phases with significant research efforts and need more domains research attention in the future.

This research paper provides future research perspective of big data applications in software engineering. It will help potential researchers to identify promising as well as underexplored application domains and focus to develop better big data applications using specific software engineering methodologies. More research in these area motivate the big data application developers and project managers to contribute more time, effort and resources in different software development life cycle (SDLC) phases of big data application development.

Future work encompasses widening the net to look for more papers. The search can be wide for more promising application domains by utilizing big data applications. The next step is to supplement the existing manual search with an automated search to make this review reproducible that will help in widening the range covered by the search to get more relevant results and avoiding false positives.

References

[1] P. Pakkla and D. Pakkala "Reference architecture and classification of technologies, products and services for big data systems" *Big Data Research*, July 2016

[2] A. Lahcen and S. Belfkih "An overview of big data opportunities, applications and tools" in *Intelligent Systems and Computer Vision*, 2015

[3] M. Vanauer, C. Bohle and B. Hellingrath "Guiding the introduction of big data in organizations: A methodology with business and data-driven ideation and enterprise architecture management based implementation" in *System Sciences, Hawaii Intl. Conf. on*. IEEE, Jan 2015

[4] M. Kramer and I. Senner "A modular software architecture for processing of big geospatial data in the cloud" *Computers & Graphics*, 2015

[5] G. Chen and Y. Wang "The evolvement of big data systems: From the perspective of an information security application" *Big Data Research*, January 2016

[6] C. Esposito, F. Palmieri and A. Castiglione "A knowledge-based platform for big data analytics based on publish/subscribe services and stream processing" *Knowledge-Based Systems*, November 2015

[7] A. Vinay, V. S. Shekhar, T. Aggrawal and S. Natarajan "Cloud based big data analytics framework for face recognition in social networks using machine learning" *Procedia Computer Science*, 2013

[8] X. Zhang and J. Chen "KASR: A Keyword-Aware Service Recommendation method on MapReduce for big data applications" *Parallel and Distributed Systems, IEEE Transactions*, 2012

[9] S. Saydam and V. V. Zira "Advancing big data for humanitarian needs" *Procedia Engineering*, 2014

[10] P. Sullivan, G. Thompson and A. Clifford "Applying data models to big data architectures" *IBM Journal of Research and Development*, 2014

[11] A. Cuzzocrea and B. Oliveira "Modeling and supporting ETL processes via a pattern-oriented and task-reusable framework" in *Tools with Artificial Intelligence, Intl. Conf*. IEEE, 2014

[12] J. Chen and R. Huang "WaaS: Wisdom as a Service" *Intelligent Systems, IEEE*, 2012

[13] C. Ordonez and S. Cabrera "Extending ER models to capture database transformations to build data sets for data mining" *Data & Knowledge Engineering*, 2014

[14] G. Casale, F. Barbier, E. D. Nitto, C. Joubert, J. Merseguer, J. F. Perez, M. Rossi and C. Sheridan "DICE: Quality-driven development of data-intensive cloud

International Journal of Scientific & Engineering Research, Volume 8, Issue Ⅲ, ÜÎÜÚÛ-2017
ISSN 2229-5518

applications" in *Proceedings of the 7th Intl. Workshop on Modeling in Software Engineering*. IEEE Press, 2015

[15] V. Kulkarni "Early experience with model-driven development of Map Reduce based big data application" in *21st Asia-Pacific Software Engineering Conference*, 2014

[16] C. Douglas "An open framework for dynamic big-data-driven Application Systems development" *Procedia Computer Science*, 2014

[17] H. Demirkan and D. Delen "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud" *Decision Support Systems*, 2013

[18] F. Yang Turner and L. Lau "A model-driven prototype evaluation to elicit requirements for a sense making support tool" in *Software Engineering Conf., Asia-Pacific*. IEEE, 2012

[19] D. Breuker "Towards model-driven engineering for big data analytics - An exploratory analysis of domain-specific languages for machine learning" in *System Sciences, Hawaii Intl. Conf. on*. IEEE, 2014

[20] V. Dmitriyev, M. Abilov, and J. M. Gomez "ELTA: New approach in designing business intelligence solutions in era of big data" *Procedia Technology*, 2014

[21] B. Hendradjaya and W. D. Sunindyo "Modeling the requirements for big data application using goal oriented approach" in *Data and Software Engineering, Conf on*. IEEE, 2014

[22] L. Huang and L. Chen "Breeze graph grammar: A graph grammar approach for modeling the software architecture of big data-oriented software systems" *Software: Practice and Experience*, 2014

[23] J. Dajda and G. Dobrowolski "Architecture dedicated to data integration" in *Intelligent Information and Database Systems*, Springer, 2015

[24] R. Girardi and L. Marinho "A domain model of web recommender systems based on usage mining and collaborative filtering" *Requirements Engineering*, 2007

[25] Y. Jararweh, M. Jarrah and Z. Alshara "CloudExp: A comprehensive cloud computing experimental framework" *Simulation Modeling Practice and Theory*, 2014

[26] D. N. Jutla, P. Bodorik and S. Ali "Engineering privacy for big data apps with the Unified Modeling Language" in *Big Data, Intl. Congress on*. IEEE, 2013

[27] H. M. Chen, R. Kazman, and O. Hrytsay "Big data system development: An embedded case study with a global outsourcing firm" in *Proceedings of the First Intl. Workshop on BIG Data Software Engineering*. IEEE Press, 2015

[28] M. Sneed and K. Erdoes "Testing big data: Assuring the quality of large databases" in *Software Testing, Verification and Validation Workshops, 8th Intl. Conf. on*. IEEE, 2015

[29] M. Jimenez, S. Mosser and M. Riveill "An architecture to support the collection of big data in the Internet of Things" in *Services, World Congress on,* IEEE, 2014

[30] K. M. Anderson "Embrace the challenges: Software engineering in a big data world" in *Proceedings of the First Intl. Workshop on BIG Data Software Engineering*. IEEE Press, 2015

[31] I. Gorton and J. Klein "Distribution, Data and Deployment: Software architecture convergence in big data systems" *IEEE Software*, vol. 32, no. 3, pp. 78–85, May 2015

[32] A. Zimmermann, I. Petrov and E. El-Sheikh "Towards service-oriented enterprise architectures for big data applications in the cloud" in *Enterprise Distributed Object Computing Conference Workshops, Intl.* IEEE, 2013

[33] M. Arias and D. Fernandez "The SOLID architecture for real-time management of big semantic data" *Future Generation Computer Systems*, 2015

[34] L. Truong and S. Dustdar "Sustainability data and analytics in cloud-based M2M systems" in *Big Data and Internet of Things: A Roadmap for Smart Environments*. Springer, 2014

[35] F. Bonomi, P. Natarajan and J. Zhu "Fog computing: A platform for Internet of Things and analytics" in *Big Data and Internet of Things: A Roadmap for Smart Environments*. Springer, 2014

[36] C. A. Boone, J. D. Ezell and L. A. Jones-Farmer "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications" *Intl. Journal of Production Economics*, 2014

[37] T. Palpanas and E. D. Valle "Towards mega-modeling: A walk through data analysis experiences" *ACM SIGMOD Record*, 2013

[38] V. Chang and M. Ramachandran "A proposed case for the cloud software engineering in security" in *Proceedings of the Intl. Workshop on Emerging Software as a Service and Analytics*, Tech. Press, 2014

[39] A. Escalona, Y. Guo and J. Offutt "A scalable big data test framework" in *Software Testing, Verification and Validation, Intl. Conf. on* IEEE, 2015

[40] M. Villari, A. Celesti and A. Puliafito architecture for the management of smart environments in IoT" in *Smart Computing Workshops, Intl. Conf. on* IEEE, 2014

[41] L. M. Pham, D. Donsez, Y. Gibello and N. de Palma "An adaptable framework to deploy complex applications onto multi-cloud platforms" in *Computing & Communication Technologies Research, Innovation and Vision for the Future, Intl. Conf. on* IEEE, 2015

[42] A. R. Najeeb and A. H. Hashim "Big data analysis solutions using MapReduce framework" in *Computer and Communication Engineering, Intl. Conf. on* IEEE, 2014

[43] H. Hata and K. Matsumoto "Bugarium: 3D interaction

International Journal of Scientific & Engineering Research, Volume 8, Issue ll, ÜÎÜÚÛ-2017
ISSN 2229-5518

for supporting large-scale bug repositories analysis" in *Companion Proceedings of the Intl. Conf. on Software Engineering*. ACM, 2014, pp. 500–503

[44] E. Begoli and J. Horey "Design principles for effective knowledge discovery from big data" in *Software Architecture and European Conference on Software Architecture, Joint Working IEEE/IFIP Conf. on*, 2012

[45] R. Lee, S. Zhang, C. H. Xia and X. Zhang "DOT: A matrix model for analyzing, optimizing and deploying software for big data analytics in distributed systems" in *Proceedings of the ACM Symposium on Cloud Computing*, 2011

[46] M. Camilli "Formal verification problems in a big data world: Towards a mighty synergy" in *Companion Proceedings of the Intl. Conf. on Software Engineering*. ACM, 2014

[47] F. Shull "Getting an intuition for big data" *Software, IEEE*, vol. 30, no. 4, pp. 3 - 6, 2013

[48] S. Bazargani, J. Brinkley and N. Tabrizi "Implement-ing conceptual search capability in a cloud-based feed aggregator" in *Innovative Computing Technology, Intl. Conf. on* IEEE, 2013

[49] W. Guo and Y. Jin "Store, schedule and switch - A new data delivery model in the big data era" in *Transparent Optical Networks, Intl. Conf. on* IEEE, 2013

[50] S. Shukla and G. Sadashivappa "A distributed randomization framework for privacy preservation in big data" in *IT in Business, Industry and Government, Conference on* IEEE, 2014

[51] Z. Liu "Research of performance test technology for big data applications" in *Information and Automation, IEEE Intl. Conf. on*, 2014

[52] C. Wang and X. Zhou "SODA: software defined accelerators for big data" in *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition,* EDA Consortium, 2015

[53] Al Zamil and S. Samarah "The application of semantic-based classification on big data" in *Information and Communication Systems, Intl. Conf. on* IEEE, 2014

[54] R. Agrawal, A. Imran and J. Walker "A layer based architecture for provenance in big data" in *Big Data, IEEE Intl. Conf. on*, 2014

[55] N. H. Madhavji "Big picture of big data software engineering: With example research challenges" in *Proceedings of the Intl. Workshop on BIG Data Software Engineering*. IEEE Press, 2015

[56] K. Kanoun, M. Ruggiero and M. V. Schaar "Low power and scalable many-core architecture for big-data stream computing" in *VLSI, IEEE Computer Society Annual Symposium on*, 2014

[57] V. Kantere, A. Nanos and N. Koziris "I/O performance modeling for big data applications over cloud infrastructures" in *Cloud Engineering, IEEE Intl. Conf. on*, 2015

[58] K. Wang and G. Chen "Breaking the boundary for whole-system performance optimization of big data" in *Proceedings of the Intl. Symposium on Low Power Electronics and Design*, IEEE Press, 2013

[59] Siva kumar and K. Kannan "Enrichment patterns for big data" in *Big Data, IEEE Intl.* 2014

[60] I. Gorton and J. Klein "Architecture knowledge for evaluating scalable databases" DTIC Document, Tech. Rep., 2015

[61] J. Han, Y. Wang and L. Liu "Big Data-as-a-Service: Definition and architecture" in *Communication Technology, IEEE Intl. Conf. on*, 2013

[62] M. Li and A. R. Butt "GERBIL: MPI + YARN" in *Cluster, Cloud and Grid Computing, IEEE/ACM Intl. Symposium on*, 2015

[63] N. Mishra "A cognitive oriented framework for IoT big-data management prospective" in *Communication Problem-Solving, IEEE Intl. Conf. on*. IEEE, 2014

[64] E. Durham, A. Rosen and R. Harrison "A model architecture for big data applications using relational databases" in *Big Data, IEEE Intl. Conf. on*, 2013

[65] A. Kashlev and S. Lu "A big data modeling methodology for Apache Cassandra" in *Big Data, IEEE Intl. Congress on*, 2015

[66] R. J. Nowling and J. Vyas "A domain-driven, generative data model for big pet store" in *Big Data and Cloud Computing, IEEE Intl. Conf. on*, 2014

[67] O. Fratu "Big data search for environmental telemetry" in *Communications and Networking, IEEE Intl. Black Sea Conf. on*, 2014

[68] A. Desai and K. Nagegowda "Advanced control distributed processing architecture (ACDPA) using SDN and Hadoop for identifying the flow characteristics and setting the quality of service (QoS) in the network" in *Advance Computing Conf., IEEE Intl.*, 2015

[69] M. Rehman, C. Liew and T. Y. Wah "UniMiner: Towards a unified framework for data mining" in *Information and Communication Technologies, Fourth World Congress on*. IEEE, 2014

[70] D. Tracey and C. Sreenan "A holistic architecture for the internet of things, sensing services and big data" in *Cluster, Cloud and Grid Computing, IEEE/ACM Intl. Symposium on*, 2013

[71] G. Vafiadis and T. Varvarigou "Enabling proactive data management in virtualized hadoop clusters based on predicted data activity patterns" in *P2P, Parallel, Grid, Cloud and Internet Computing, Eighth Intl. Conf. on*. IEEE, 2013

[72] B. T. Kumara, J. Zhang and R. Koswatte "Ontology based workflow generation for intelligent big data analytics" in *Web Services, IEEE Intl. Conf. on*, 2015

International Journal of Scientific & Engineering Research, Volume 8, Issue ÙÙ ÜÎÜÚÛ-2017
ISSN 2229-5518

[73] U. Hedlund and K. M. Bjork "A generalized scalable software architecture for analyzing temporally structured big data in the cloud" in *New Perspectives in Information Systems and Technologies, Volume 1*. Springer, 2014

[74] H. S. Lamba and S. K. Dubey "Analysis of requirements for big data adoption to maximize IT business value" in *Reliability, Infocom Technologies and Optimization (Trends and Future Directions), 2015 4th International Conference on*, 2015

[75] H. M. Chen and S. Haziyev "Strategic prototyping for developing big data systems" *IEEE Software*, 2016

[76] W. Zhang and Y. Liu "A deep-intelligence framework for online video processing" *IEEE Software*, 2016

[77] F. Marconi, P. Jamshidi and A. Nodari "Continuous architecting of stream-based systems" in *13th Working IEEE/IFIP Conf. on Software Architecture*, 2016

[78] A. Zimmermann, R. Schmidt and E. El-Sheikh "Adaptable enterprise architectures for software evolution of SmartLife ecosystems" in *Intl. Enterprise Distributed Object Computing Conf. Workshops and Demonstrations*. IEEE, 2014

[79] M. Grechanik and D. Poshyvanyk "Sanitizing and minimizing databases for software application test outsourcing" in *Intl. Conf. on Software Testing, Verification and Validation*, IEEE, 2014

[80] T. C. Chiang and Y. C. Zeng "Nonparametric discovery of contexts and preferences in smart home environments" in *Systems, Man, and Cybernetics, IEEE Intl. Conf. on*, 2015

[81] S. Anuradha "State of art in testing for big data" in *IEEE Intl. Conf. on Computational Intelligence and Computing Research*, 2015

[82] M. O. Gokalp and P. E. Eren "A cloud based architecture for distributed real time processing of continuous queries" in *Conf. on Software Engineering and Advanced Applications*, 2015

[83] L. Zhu, D. Sun and Q. Lu "Building pipelines for heterogeneous execution environments for big data processing" *IEEE Software*, 2016

[84] L. JunYi "Big data transformation testing based on data reverse engineering" in *Intl. Conf. on Ubiquitous Intelligence and Computing, Intl Conf.*, 2015

[85] K. Taneja, Q. Zhu, D. Duggan, and T. Tung, "Linked enterprise data model and its use in real time analytics and context-driven data discovery," in *IEEE Intl. Conf. on Mobile Services*, 2015

[86] Z. Liu, "Research of performance test technology for big data applications," in *Information and Automation, IEEE Intl. Conf. on*, 2014

[87] H. Zhou, H. Zhang and H. Lin "An empirical study on quality issues of production big data platform" in *IEEE/ACM Intl. Conf. on Software Engineering*, 2015

[88] A. Samuel, H. Haseeb and S. Basalamah "A framework for composition and enforcement of privacy-aware and context-driven authorization mechanism for multimedia big data" *IEEE Transactions on Multimedia*, 2015

[89] A. Doyle, G. Katz, K. Summers, C. Ackermann, Z. Lim, L. Zhao, P. Butler and N. Rama krishnan "The EMBERS architecture for streaming predictive analytics" in *Big Data, IEEE Intl. Conf. on*, 2014

[90] F. Shen "A pervasive framework for real-time activity patterns of mobile users" in *Pervasive Computing and Communication Workshops, IEEE Intl. Conf. on*, 2015

[91] S. Yang and J. Wang "An automatic discovery framework of cross-source data inconsistency for web big data" in *Intl. Conf. on Advanced Cloud and Big Data*, 2015

[92] H. M. Chen, R. Kazman and S. Haziyev "Agile big data analytics for web-based systems: An architecture-centric approach" *IEEE Transactions on Big Data*, 2016

[93] S. Singh and Y. Liu "A cloud service architecture for analyzing big monitoring data" *Tsinghua Science and Technology*, pp. 55–70, 2016

[94] J. M. Smith and A. Mockus "Exploring a framework for identity and attribute linking across heterogeneous data systems" in *Proceedings of the 2nd Intl.* ACM, 2016

[95] P. Nair and V. Saravagi "Model-driven observability for big data storage" in *13th Working IEEE/IFIP Conf. on Software Architecture*, 2016

[96] D. Safaric and J. Visser "Streaming soft-ware analytics" in *Proceedings of the 2nd Intl. Work-shop on BIG Data Software Engineering*. ACM, 2016

[97] S. Tajfar and D. A. Tamburri "Towards a model-driven design tool for big data architectures" in *Proceedings of the 2nd Intl. Workshop on BIG Data Software Engineering*, ACM, 2016

[98] I. Noorwali and N. H. Madhavji "Under-standing quality requirements in the context of big data systems" in *Proceedings of the 2nd Intl. Workshop on BIG Data Software Engineering*. ACM, 2016

[99] R. State and T. Engel "A big data architecture for large scale security monitoring" in *IEEE Intl. Congress on Big Data*, June 2014

[100] F. Zhang, S. U. Khan and K. Hwang "A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications" *Future Generation Computer Systems*, 2015

[101] M. Zhang, C. Liu and N. Rishe "Terra Fly Geo cloud: An online spatial data analysis and visualization system"*ACM Transactions on Intelligent System and Technology*, 2015

[102] Z. Xu, Y. Liu and L. Chen "Knowledge: A semantic link network based system for organizing large scale online news events" *Future Generation Computer Systems*, 2015

International Journal of Scientific & Engineering Research, Volume 8, Issue Ⅲ, ÜÎÜÚÛ-2017
ISSN 2229-5518

[103] T. Shah, F. Rabhi and P. Ray "Investigating an ontology-based approach for big data analysis of inter-dependent medical and oral health conditions" *Cluster Computing*, 2014

[104] Q. Yao, Y. Tian and Y. Qian "Design and Development of a Medical Big Data Processing System based on Hadoop" *Journal of medical systems*, 2015

[105] M. A. Saleem and S. Lee "Trajectory patterns mining towards life care provisioning" *Wireless Personal Communications*, 2014

[106] D. Bromley and V. Daggett "DIVE: A graph-based visual-analytics framework for big data" *IEEE Computer Graphics and Applications*, 2014

[107] A. Naseer and E. Waraich "A big data approach for proactive healthcare monitoring of chronic patients" in *Intl. Conf. on Ubiquitous and Future Networks (ICUFN)*, 2016

[108] K. Gai, L. C. Chen and M. Liu "Electronic health record error prevention approach using ontology in big data" in *High Performance Computing and Communications, Intl. Symposium on Cyberspace Safety and Security, Intl. Conf. on Embedded Software and Systems, Intl. Conf. on*. IEEE, 2015

[109] D. Zhang and X. H. Hu "A framework for ensuring the quality of a big data service" in *IEEE Intl. Conf. on Services Computing*, 2016

[110] R. Sendelj and W. Hack "Towards semantically enabled development of service-oriented architectures for integration of socio-medical data" in *Mediterranean Conference on Embedded Computing*, 2016

[111] K. Kaur and R. Rani "A smart polyglot solution for big data in healthcare" *IT Professional*, 2015

[112] A. Khalil and Z. Tari "BDCaM: Big data for context-aware monitoring - A Personalized Knowledge Discovery Framework for Assisted Health-care" *IEEE Transactions on Cloud Computing*, 2015

[113] R. Giachetta "A framework for processing large scale geospatial and remote sensing data in MapReduce environment" *Computers & Graphics*, 2015

[114] M. Muller and D. Kadner "Moving code-sharing geo processing logic on the web" *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013

[115] T. Shelton, M. Graham and M. Zook "Mapping the data shadows of Hurricane Sandy: Un-covering the sociospatial dimensions of big data" *Geo-forum*, 2014

[116] A. Padmanabhan, Z. Zhang and K. Soltani "A scalable framework for spatiotemporal analysis of location-based social media data" *Computers, Environment and Urban Systems*, 2015

[117] M. Akmal "Assembling cloud-based geographic information systems: A pragmatic approach using off-the-shelf components" *Cloud Computing with e-Science Applications*, 2015

[118] J. Alder and S. Hostetler "Web based visualization of large climate data sets" *Environmental Modelling & Software*, 2015

[119] M. Deng "Building open environments to meet big data challenges in Earth sciences" in *Big Data: Techniques and Technologies in Geo-informatics*. CRC Press, 2013

[120] S. Fang and Z. Liu "An integrated system for regional environmental monitoring and management based on internet of things" *Industrial Informatics, IEEE Transactions*, 2014

[121] I. Manssour and L. G. Fernandes "Towards a domain-specific language for geospatial data visualization maps with big data sets" in *IEEE/ACS Intl. Conf. of Computer Systems and Applications*, 2015

[122] J. Anderson, R. Soden and L. Palen "EPIC-OSM: A software framework for OpenStreetMap Data Analytics" in *Hawaii Intl. Conf. on System Sciences*, 2016

[123] N. Pelekis and D. Janssens "On the management and analysis of our life steps" *SIGKDD Explorer, News*, 2014

[124] Y. Zhang, M. Chen and V. Leung "CAP: Community Activity Prediction based on big data analysis" *Network, IEEE*, 2014

[125] T. Hong "Occupancy schedules learning process through a data mining framework" *Energy and Buildings*, 2015

[126] R. Sinnott and M. Tomko "Modeling coordinated multiple views of heterogeneous data cubes for urban visual analytics" *Intl. Journal of Digital Earth*, 2014

[127] D. Bonino and G. Procaccianti "Exploiting semantic technologies in smart environments and grids: Emerging roles and case studies" *Science of Computer Programming*, 2014

[128] C. Dobre and F. Xhafa "Intelligent services for big data science" *Future Generation Computer Systems*, 2014

[129] X. Zhang and H. B. Lim "Smart traffic cloud: An infrastructure for traffic applications" in *Parallel and Distributed Systems, Intl. Conf. on*. IEEE, 2012

[130] P. A. Mathew and T. Walter "Big-data for building energy performance: Lessons from assembling a very large national database of building energy use" *Applied Energy*, 2015

[131] X. Zhang and H. B. Lim "A cooperative sensing and mining system for transportation activity survey" in *IEEE Wireless Communications and Networking Conference*, 2014

[132] Y. Zhu, J. Zhang and Y. Chen "Improving power grid monitoring data quality: An efficient machine learning framework for missing data prediction" in *Intl. Conf. on Embedded Software and Systems, Intl. Conf. on*. IEEE, 2015

International Journal of Scientific & Engineering Research, Volume 8, Issue ⅢⅠ, ÜÎÜÚÛ-2017
ISSN 2229-5518

[133] B. Cheng, S. Longo and M. Bauer "Building a big data platform for smart cities: Experience and lessons from Santander" in *IEEE Intl. Congress on Big Data*, June 2015

[134] D. Parikh and A. Hampapur "Improving rail network velocity: A machine learning approach to predictive maintenance" *Transportation Research Part C: Emerging Technologies*, 2014

[135] A. Thaduri and U. Kumar "Railway assets: A potential domain for big data analytics," *Procedia Computer Science*, 2015

[136] C. D. Cottrill and S. Derrible "Leveraging big data for the development of transport sustainability indicators" *Journal of Urban Technology*, 2015

[137] S. D. Martino and T. Rustemeyer "An architecture to process massive vehicular traffic data" in *Intl. Conf. on P2P, Parallel, Grid, Cloud and Internet Computing*, 2015

[138] L. Morandini "SMASH: A cloud-based traffic data architecture for big data processing and visualization" in *IEEE Intl. Conf. on Data Science and Data Intensive Systems*, 2015

[139] J. Yang "A big-data processing framework for uncertainties in transportation data" in *Fuzzy Systems, IEEE Intl. Conf. on*, Aug 2015

[140] G. Menon and V. Ramamurthy "Intelligent operational dash boards for smarter commerce using big data" *IBM Journal of Research and Development*, 2014

[141] T. Yang and S. Xu "Context-based ontology-driven recommendation strategies for tourism in ubiquitous computing" *Wireless Personal Communications*, 2014

[142] R. P. Kinsley and J. Portenoy "Perspectives of emerging museum professionals on the role of big data in museums" in *System Sciences, Hawaii Intl. Conf. on* IEEE, 2015

[143] L. Deng and J. Gao "Building a big data analytics service framework for mobile advertising and marketing" in *Big Data Computing Service and Applications, IEEE Intl. Conf. on*, 2015

[144] G. Todoran and A. Apostu "Cloud computing for extracting price knowledge from big data" in *Complex, Intelligent, and Software Intensive Systems, Intl. Conf. on*. IEEE, 2015

[145] Q. Huang and C. Xu "A data-driven framework for archiving and exploring social media data" *Annals of GIS*, 2014

[146] A. Immonen and E. Ovaska "Evaluating the quality of social media data in big data architecture," *IEEE Access*, 2015

[147] S. Ahamed "A reference architecture for social media intelligence applications in the cloud" in *Computer Software and Applications Conference*, IEEE, 2015

[148] S. Bristol "SemantEco: A semantically powered modular architecture for integrating distributed environmental and ecological data" *Future Generation Computer Systems*, 2014

[149] C. Beal and J. Flynn "Toward the digital water age: Survey and case studies of Australian water utility smart metering programs" *Utilities Policy*, 2015

[150] E. Moguel and J. Hernandez "Multilayer big data architecture for remote sensing in eolic parks" *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015

[151] D. Dutta and I. Bose "Managing a big data project: The case of Ramco Cements limited" *Intl. Journal of Production Economics*, 2015

[152] N. Kushiro and K. Takahara "Model oriented system design on big-data" *Procedia Computer Science*, 2014

[153] J. Lee and S. Yang "Service innovation and smart analytics for industry 4.0 and big data environment" *Procedia CIRP*, 2014

[154] Y. Liu and H. Zhang "Research and application of one key publishing technologies for meteorological service products" in *IEEE Intl. Conf. on Big Data Analysis*, 2016

[155] J. Caldera "Towards a data processing architecture for the weather radar of the INTA Anguil" in *Intl. Workshop on Data Mining with Industrial Applications*, 2015

[156] L. Zhang "A framework to model big data driven complex cyber physical control systems" in *Automation and Computing, Intl. Conf. on* IEEE, 2014

[157] S. Mosser "Software development support for shared sensing infrastructures: A generative and dynamic approach" in *Software Reuse for Dynamic Systems in the Cloud and Beyond*. Springer, 2014

[158] L. Zhang "Designing big data driven cyber physical systems based on AADL" in *Systems, Man and Cybernetics, IEEE Intl. Conf. on*, 2014

[159] M. Rahmes and C. Casseus "Multi-disciplinary ontological geo-analytical incident modeling" in *IEEE Consumer Communications and Networking Conference*, 2015

[160] J. Klein and K. Cooper "A reference architecture for big data systems in the National Security domain" in *Proceedings of Big Data Software Engineering Workshop on*. ACM, 2016

[161] M. Andrade and A. Sefidcon "The use of distributed processing and cloud computing in agricultural decision making support systems" in *Cloud Computing, Intl. Conf. on* IEEE, 2014

[162] R. Dutta and A. Das "Development of an intelligent environmental knowledge system for sustainable agricultural decision support" *Environmental Modeling & Software*, 2014

[163] N. Sun and J. Morris "iCARE: A framework for big data-based banking customer analytics" *IBM Journal of*

International Journal of Scientific & Engineering Research, Volume 8, Issue ⅡⅠ, ÜÎÜÚÚⅡ-2017
ISSN 2229-5518

Research and Development, 2014

[164] A. Chaparro "Using the web to monitor a customized unified financial portfolio" in *Advances in Conceptual Modeling*, Springer, 2012

[165] P. Chen and L. Xu "Research on Warship Communication Operation and Maintenance Management Based on Big Data" in *Cloud Computing and Big Data, Intl. Conf.* IEEE, 2014

[166] S. Riggi and Krokos "Towards a big data exploration framework for astronomical archives" in *High Performance Computing & Simulation, Intl. Conf. on* IEEE, 2014

[167] E. Begoli "A short survey on the state of the art in architectures and platforms for large scale data analysis and knowledge discovery from data" in *Proceedings of the WICSA/ECSA Companion Volume,* ACM, 2012

[168] M. Chen and Y. Liu "Big data: A survey" *Mobile Networks and Applications*, 2014

[169] C. P. Chen and C. Y. Zhang "Data-intensive applications, challenges, techniques and technologies: A survey on big data" *Information Sciences*, 2014

[170] J. S. Ward and A. Barker "Undefined by data: A survey of big data definitions" *arXiv preprint arXiv-1309*, 2013

[171] A. B. Bener and A. Mockus "Software Engineering for Big Data Systems" *IEEE Software*, 2016

[172] K. S. Yim "Norming to Performing: Failure analysis and deployment automation of big data software developed by highly iterative models" in *Software Reliability Engineering, Intl. Symposium on*. IEEE, 2014

[173] C. E. Cuesta and J. D. Fernandez "Towards an architecture for managing big semantic data in real-time" in *Software Architecture*, Springer, 2013

[174] R. S. Pressman "*Software Engineering: A Practitioner's Approach"* Palgrave Macmillan, 2005

[175] J. S. Saltz and I. Shamshurin "Exploring the process of doing data science via an ethnographic study of a media advertising company" in *Big Data, IEEE Intl. Conf. on*, 2015