

# Software Engineering for Big Data Projects: Domains, Methodologies and Gaps

Vijay Dipti Kumar

David R. Cheriton School of Computer Science  
University of Waterloo  
Waterloo, Canada  
vdkumar@uwaterloo.ca

Paulo Alencar

David R. Cheriton School of Computer Science  
University of Waterloo  
Waterloo, Canada  
palencar@cs.uwaterloo.ca

**Abstract—Context:** Big data has become the new buzzword in the information and communication technology industry. Researchers and major corporations are looking into big data applications to extract the maximum value from the data available to them. However, developing and maintaining stable and scalable big data applications is still a distant milestone. **Objective:** To look at existing research on how software engineering concepts, namely the phases of the software development project life cycle (SDPLC), can help build better big data application projects. **Method:** A literature survey was performed. A manual search covered papers returned by search engines resulting in approximately 2,000 papers being searched and 170 papers selected for review. **Results:** The search results helped in identifying data rich application projects that have the potential to utilize big data successfully. The review helped in exploring SDPLC phases in the context of big data applications and performing a gap analysis of the phases that have yet to see detailed research efforts but deserve attention.

**Keywords—**Big data, software engineering, application projects, domains, methodologies, gaps, literature review.

## I. INTRODUCTION

There have been several success stories of big data being used by technology giants to dominate their competitors in areas such as social media, search engines, e-commerce and video streaming services. Few of the popular players and leaders in these fields leveraging big data are Facebook, LinkedIn, Twitter, Google, Amazon, and Netflix. The success of these big data use cases have piqued the interest of and spurred numerous companies to take an interest in developing big data applications, to extract maximum value from all the data that is becoming available.

According to a Gartner survey<sup>1</sup>, 64% out of 720 respondents of the survey had invested or planned to invest in big data applications in 2013. However, less than 8% had actually deployed at the time of the survey. As shown in a case study of optimizing the manufacturing process of digital displays [1], it is possible to enhance process efficiency using big data. Different sectors such as healthcare, trading, agriculture, tourism and politics, in which big data is being used to stakeholders' advantage, are detailed in [2]. Enriching customer experience using big data has been exemplified by the "People you may know" option offered by Facebook and LinkedIn, movie recommendations by Netflix, and even the

"Customers Who Bought This Item Also Bought" service provided by Amazon.

### A. Big Data and Software Engineering

Big data has been characterised by 4Vs: Volume, Velocity, Variety and Veracity. Volume implies the data explosion that has come to mark the last decade in the world of computing. Velocity is the constraint that demands real time processing of the data available, the failure of which would result in its loss or obsolescence. Variety means data format could be structured, semi-structured or unstructured or have multiple sources and applications must be capable of handling these sources. Finally, veracity implies that the available data, historical or even real-time streaming data, might have to be cleaned prior to processing to ensure its usefulness to the application.

The whole process of developing software is fraught with errors which arise due to changes in requirements, the environment or due to communication issues between the stakeholders. This is true for software development for ordinary applications even today despite the fact we have been developing software for more than 20 years. When factors such as the 4Vs are taken into account, the complexities involved in developing software for big data application projects only increase.

Despite the complexities involved in following the SDPLC phases in building a big data application, these phases and concepts must be leveraged to build robust and scalable big data applications projects. It would be to the advantage of the stakeholders and developers involved in building a big data application that the best practices and methodologies laid down by the software engineering research community be applied to build systems that are fault tolerant and capable of handling even more data than envisioned at the time of its creation.

### B. Motivation

There have been numerous reviews on big data itself. One discusses state of the art in architecture and large scale data analysis platforms [168], one is a comprehensive big data survey [169], another describes the related technologies and the acquisition and applications of big data [170] and one even lists the different definitions of big data [171]. The guest editor's introduction to one of the issues of the IEEE Software magazine [172] discussed the software engineering challenges in building data-intensive, or big data software

<sup>1</sup>Big Data Gartner survey-<http://www.gartner.com/newsroom/id/2593815>

TABLE I. APPLICATION DOMAIN CATEGORIES &amp; CLASSIFICATION

Application Domain	Papers	Count
Information Technology	[1], [3–99]	98
Healthcare	[100–112]	13
Geospatial Data Processing/Geographic Information Systems	[4], [101], [113–122]	12
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
Environmental Monitoring/Conservation	[4], [85], [120], [149–151]	6
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

systems. However, no comprehensive study reviewing existing software engineering research methodologies for enabling development of big data application project was found.

### C. Goals and Contributions

The main goal of this study was to look into the existing research on applying software engineering to big data application projects. The results from this literature survey help identify the domains that have been studied in detail for big data application development. It also helped recognize the SDPLC phases that were most commonly being utilized in developing big data applications. Additional results include pinpointing the domains that have the potential to deploy big data application projects with advantageous outcomes and have seen early research efforts in the field but remain underexplored and definitely deserve more attention from researchers.

## II. RESEARCH METHOD

For conducting the literature review, popular academic search engines, namely Scopus, Web of Science and IEEE Xplore Digital Library were targeted. The search was performed using the Command option under Advanced Search section of these search engines. A combination of keywords related to SDPLC phases were selected from the software engineering standard textbook *Software Engineering: A Practitioner’s Approach* [175] to understand which subfields were popular among researchers.

In the context of software engineering, the search terms used were architecture, evolution, process, quality, reuse, specification, requirement(s) engineering, design, “domain mod-

eling”, testing, verification, validation, maintenance, quality, analysis, framework, process, and patterns.

The idea behind formulating the queries was to search for papers that combined topics about software engineering and big data. We have adopted an extensive search and selection process to identify a set of studies that is as complete as possible. Our search process involved automatic and manual searching. An example of the pattern of queries used in our study is as follows:

“big data” AND (engineering OR requirement OR specification OR design OR architecture OR analysis OR testing OR verification OR validation OR maintenance OR framework OR quality OR design OR evolution OR patterns OR process OR reuse OR “domain modeling”)

The abstract, introduction and conclusion of each search result was examined in detail to ensure the papers were eligible to be included in the study. Reproducibility was not given a priority because a manual search was the main method applied.

Admittedly, the quality of our results depends on the search queries used and the efficiency of the search engines. Numerous publications from relevant conferences related to big data did not show up in the search results because of the SDPLC specific search terms used in the queries. For example, the use of big data in the field of advertising [176] was not covered because there is no specific SDPLC phase referenced or used in this study despite the fact that the paper dealt with an interesting application domain, namely advertising.

TABLE II. SDPLC PHASES

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

### A. Research Questions

The main research questions that were addressed through the study were:

- RQ1. Which application domains have received attention for the development of big data application projects and which domains require more attention?
- RQ2. Which SDPLC phases were used to enable big data applications and which fields need more research efforts?

### B. Classification Criteria

The following were the main criteria taken into account when analyzing and categorizing each paper. Under each criteria, multiple categories were identified. The first criteria categorized the papers on the basis of the application domain addressed by the paper and the second criteria identified the SDPLC phase studied.

- C1. Which application domain does the paper belong to?
- C2. Which SDPLC phase is utilized in the paper?

## III. LIMITATIONS

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

Admittedly, there is room for improvement and opportunity to include more relevant papers. More papers may have been added to the these search engines after the queries for this study were run. Any relevant papers so missed would potentially be due to a matter of timing rather than oversight.

## IV. RESULTS

The main purpose of this literature review in addition to looking into the existing research of applying software engineering to big data, was to perform a gap analysis of the research till date. The gap analysis helped identify the application domains and the SDPLC phases in the context of big data that have not yet received much attention from researchers but have huge potential.

The application domains identified through this study are illustrated in Table I and the classification of the SDPLC phases for big data addressed in each paper is listed 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?

Table I lists all the application domains found by analyzing the papers in this study. The papers that proposed new methods or customized versions of existing technology but did not mention a specific domain, such as healthcare, military, or infrastructure, were classified into the “Information Technology” category. Out of the 170 papers selected in this review, the number of papers dealing with this category was 98, as can be seen in the column “Count” of Table I. Nearly 57% of the analyzed papers focused on the Information Technology domain, which indicates that most of the papers are focusing on topics that would directly affect the world of computing. The primary observation was that researchers focused on improving existing technologies to better suit current requirements.

Application domains such as healthcare and the banking and financial industry are commonly believed to be data rich. The healthcare industry is a source of huge amounts of data sourced from the electronic medical records of patients. Data from hospitals, clinics, medical governing bodies and even insurance providers can be mined to study disease affliction rates, patterns and susceptibility trends. Analyzing medical

data can also help in developing innovative treatment methods, customize more effective and economical treatment plans, or even help healthcare professionals in dispensing medication to groups of patients suffering from similar afflictions that have identical medical history of symptoms and reactions.

The banking and financial industry has access to the monetary blueprint of the world. The tremendous amount of transactional data that commercial banks handle on a daily basis can be used to better understand the spending patterns of its customers. Credit card usage, mobile banking application usage patterns, mortgage and credit history of customers can give banks greater insight into the needs of its customers and help tailor their products accordingly, thereby enhancing customer experience and satisfaction.

Other data rich domains that have seen early research are infrastructure, transport and manufacturing. Metropolitan authorities responsible for traffic management and building infrastructure facilities can use current as well as historical data to build smart cities and green buildings that use minimum amounts of energy and water for heating and maintenance.

The aviation industry and the military also have huge potential for using big data for better performance and maintenance of equipment, aircrafts and vehicles. Interestingly, Geographic Information Systems (GIS) have attracted the attention of big data researchers. This domain can enhance and benefit more from the data available from location services that come built into several mobile and Internet applications used today.

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

RQ2. Which SDPLC phases were used to enable big data applications and which fields need more research efforts?

Table II provides a breakdown of the SDPLC phases that were addressed by each paper. Numerous papers dealt with methodologies involving software architectures and frameworks and a good number dealt with design. This does not come as a surprise since building big data applications with the latest cutting edge technological developments would primarily deal with the design and architecture of a system. Thus, we see a lot of research done that specifically deals with designs and software architectures or frameworks, which act as templates for building similar software systems.

SDPLC phases such as maintenance, validation, verification and quality assurance of big data applications have not seen much research. Only two papers were identified that explicitly reported research on the topic of verification [46] [109] and were classified under "Verification/Validation" in Table II. Verification and validation processes ensure that the software system has been designed and will function according to the requirements and design foundations laid down before development.

Similarly, only two papers were found that directly dealt with "Maintenance" [55] [166]. Maintenance of the software systems is what keeps the software application running smoothly even after years of deployment. These areas are extremely important to ensure that once a big data application is

built, it continues to meet its goals and satisfy its stakeholders' expectations without compromising performance and remains scalable in the future when even more data could be leveraged.

The requirements subfield fares comparatively better, which could be expected since, although the requirements gathering for big data applications due to the 4V properties of the data involved is a complex process, application data and processing requirements need to be analyzed, to some extent, before the applications are built. Conducting requirements research can provide pointers to new adopters of big data technology, and help to establish benchmarks and standard practices.

## B. Classification Procedure

The papers were classified on the basis of the classification criteria C1 and C2 mentioned in an earlier section. Each paper was analyzed and categorized on the basis of application domain and SDPLC phase. For example, if a paper discussed the application or implementation of big data in the healthcare industry then its application domain was classified as "Healthcare". If a paper proposed a domain specific language or described an ontology technique, it was categorized under "Domain Specific Language/Ontology".

Papers that did not mention any application domain explicitly were classified under "Information Technology". Several papers provided a framework approach for their system as well as an architectural implementation. However, if the author(s) of these papers identified the contribution of the paper as a framework-oriented method then the paper was designated the category "Framework" and not "Architecture".

## V. DISCUSSION

As illustrated by the results of this review in Tables I and II, there are noticeable differences in the amount of research attention received by the different application domains and SDPLC phases. The Information Technology application domain received much more research attention when compared to other important and promising data rich domains such as Healthcare and the Banking and Financial Industry.

There is a lot of potential for making technological advances using big data in domains identified by this review such as Aviation, Infrastructure, Transport and Environmental Monitoring/Conservation. These domains and many others have not yet witnessed much work from researchers but we need to prioritize these domains. More focused research utilizing big data applications has the ability to transform the software management and development of these domains.

Research on software requirements methodologies for big data applications would help reduce the chances of system errors, project failures and unsatisfied stakeholder expectations. The technologies and environment around big data continues to evolve constantly and big data applications designed today have to deal with these unknown but inevitable changes, which makes the need for requirements research even more urgent.

Similarly, more research needs to be conducted into enhancing the existing methods and practices and developing novel methodologies for maintenance, testing, validation/verification and quality assurance of big data applications.

The high stakes and risks involved, such as system unpredictability and project failure in big data applications, can be mitigated by conducting thorough system testing, validation and verification and laying the foundations for good software quality assurance practices.

## VI. CONCLUSION & FUTURE WORK

This is mostly likely to be the first comprehensive study into existing software engineering research in context of big data applications. The purpose of this literature review was to understand how software engineering research till date was enabling big data application projects and to focus on research that highlights use of the SDPLC phases in building robust, scalable big data applications.

A gap analysis was performed to identify the more popular application domains among researchers in this field. This analysis also aimed at revealing the main SDPLC phases that have seen significant research efforts and the domains and phases that need more research attention in the future.

This paper aims to provide perspective to future researchers looking into big data applications from a software engineering point of view. It can help potential researchers identify promising but underexplored application domains and focus on using specific software engineering methodologies to develop better big data applications. More research in these areas should motivate and help big data application developers and project managers to contribute more time, effort and resources in the different SDPLC phases of big data application development.

Future work encompasses widening the net to look for more papers. There may be more research published after the time of the initial search for this study was performed. By widening the search, more promising application domains can be identified that could benefit from utilizing big data applications.

With respect to the search method, the next step is to supplement the existing manual search with an automated search to make this review reproducible by other researchers. This would also help in widening the range covered by the search to get more relevant results and avoiding false positives.

## REFERENCES

- [1] P. Pääkkönen and D. Pakkala, "Reference architecture and classification of technologies, products and services for big data systems," *Big Data Research*, 2015.
- [2] F. Z. Benjelloun, A. 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. Hellinrath, "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, 2015.
- [4] M. Krämer and I. Senner, "A modular software architecture for processing of big geospatial data in the cloud," *Computers & Graphics*, 2015.
- [5] G. Chen, S. Wu, and Y. Wang, "The evolvement of big data systems: From the perspective of an information security application," *Big Data Research*, 2015.

- [6] C. Esposito, M. Ficco, F. Palmieri, and A. Castiglione, "A knowledge-based platform for big data analytics based on publish/subscribe services and stream processing," *Knowledge-Based Systems*, 2015.
- [7] A. Vinay, V. S. Shekhar, J. Rituparna, T. Aggrawal, K. B. Murthy, and S. Natarajan, "Cloud based big data analytics framework for face recognition in social networks using machine learning," *Procedia Computer Science*, 2015.
- [8] S. Meng, W. Dou, X. Zhang, and J. Chen, "KASR: A Keyword-Aware Service Recommendation method on MapReduce for big data applications," *Parallel and Distributed Systems, IEEE Transactions on*, 2014.
- [9] S. O. Fadiya, S. Saydam, and V. V. Zira, "Advancing big data for humanitarian needs," *Procedia Engineering*, 2014.
- [10] P. O'Sullivan, G. Thompson, and A. Clifford, "Applying data models to big data architectures," *IBM Journal of Research and Development*, 2014.
- [11] O. Belo, A. Cuzzocrea, and B. Oliveira, "Modeling and supporting ETL processes via a pattern-oriented, task-reusable framework," in *Tools with Artificial Intelligence, Intl. Conf. on.* IEEE, 2014.
- [12] J. Chen, J. Ma, N. Zhong, Y. Yao, J. Liu, R. Huang, W. Li, Z. Huang, Y. Gao, and J. Cao, "WaaS: Wisdom as a Service," *Intelligent Systems, IEEE*, 2014.
- [13] C. Ordóñez, S. Maabout, D. S. Matusевич, and W. Cabrera, "Extending ER models to capture database transformations to build data sets for data mining," *Data & Knowledge Engineering*, 2014.
- [14] G. Casale, D. Ardagna, M. Artac, F. Barbier, E. D. Nitto, A. Henry, G. Iuhasz, C. Joubert, J. Merseguer, V. I. Munteanu, J. F. Pérez, D. Petcu, M. Rossi, C. Sheridan, I. Spais, and D. Vladušić, "DICE: Quality-driven development of data-intensive cloud applications," in *Proceedings of the 7th Intl. Workshop on Modeling in Software Engineering.* IEEE Press, 2015.
- [15] A. Rajbhoj, V. Kulkarni, and N. Bellarykar, "Early experience with model-driven development of MapReduce based big data application," in *21st Asia-Pacific Software Engineering Conference*, 2014.
- [16] C. C. Douglas, "An open framework for dynamic big-data-driven Application Systems (DBDDAS) 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, L. Lau, and V. Dimitrova, "A model-driven prototype evaluation to elicit requirements for a sensemaking 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] P. M. Marín-Ortega, V. Dmitriyev, M. Abilov, and J. M. Gómez, "ELTA: New approach in designing business intelligence solutions in era of big data," *Procedia Technology*, 2014.
- [21] H. Eridaputra, B. Hendradjaya, and W. D. Sunindyo, "Modeling the requirements for big data application

- using goal oriented approach,” in *Data and Software Engineering, Intl. Conf on.* IEEE, 2014.
- [22] C. Li, 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. B. Marinho, “A domain model of web recommender systems based on usage mining and collaborative filtering,” *Requirements Engineering*, 2007.
- [25] Y. Jararweh, M. Jarrah, Z. Alshara, M. N. Alsaleh, M. Al-Ayyoub *et al.*, “CloudExp: A comprehensive cloud computing experimental framework,” *Simulation Modelling 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, S. Haziyevev, 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] H. 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] C. Cecchinell, 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, Deployment: Software architecture convergence in big data systems,” *IEEE Software*, vol. 32, no. 3, pp. 78–85, May 2015.
- [32] A. Zimmermann, M. Pretz, G. Zimmermann, D. G. Firesmith, 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. A. Martínez-Prieto, C. E. Cuesta, M. Arias, and J. D. Fernández, “The SOLID architecture for real-time management of big semantic data,” *Future Generation Computer Systems*, 2015.
- [34] H.-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, R. Milito, 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] B. T. Hazen, 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] S. Ceri, T. Palpanas, E. D. Valle, D. Pedreschi, J.-C. Freytag, and R. Trasarti, “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.* SciTePress, 2014.
- [39] N. Li, 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, M. Fazio, and A. Puliafito, “Alljoyn Lambda: An architecture for the management of smart environments in IoT,” in *Smart Computing Workshops, Intl. Conf. on.* IEEE, 2014.
- [41] L. M. Pham, A. Tchana, D. Donsez, V. Zurczak, P.-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] S. B. Elagib, A. R. Najeeb, A. H. Hashim, and R. F. Olanrewaju, “Big data analysis solutions using MapReduce framework,” in *Computer and Communication Engineering, Intl. Conf. on.* IEEE, 2014.
- [43] P. Yongpisanpop, H. Hata, and K. Matsumoto, “Bugarium: 3D interaction 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] Y. Huai, 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, “Implementing conceptual search capability in a cloud-based feed aggregator,” in *Innovative Computing Technology, Intl. Conf. on.* IEEE, 2013.
- [49] W. Sun, F. Li, W. Guo, Y. Jin, and W. Hu, “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, X. Li, and X. Zhou, “SODA: software defined fpga based accelerators for big data,” in *Proceedings of*

- the Design, Automation & Test in Europe Conference & Exhibition.* EDA Consortium, 2015.
- [53] M. G. 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, C. Seay, and J. Walker, "A layer based architecture for provenance in big data," in *Big Data, IEEE Intl. Conf. on,* 2014.
- [55] N. H. Madhavji, A. Miransky, and K. Kontogiannis, "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, D. Atienza, and M. Van Der Schaar, "Low power and scalable many-core architecture for big-data stream computing," in *VLSI, IEEE Computer Society Annual Symposium on,* 2014.
- [57] I. Mytilinis, D. Tsoumakos, 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] Y. Li, K. Wang, Q. Guo, X. Li, X. Zhang, G. Chen, T. Liu, and J. Li, "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] K. Holley, G. Sivakumar, and K. Kannan, "Enrichment patterns for big data," in *Big Data, IEEE Intl. Congress on,* 2014.
- [60] I. Gorton, J. Klein, and A. Nurgaliev, "Architecture knowledge for evaluating scalable databases," DTIC Document, Tech. Rep., 2015.
- [61] E. Xinhua, J. Han, Y. Wang, and L. Liu, "Big Data-as-a-Service: Definition and architecture," in *Communication Technology, IEEE Intl. Conf. on,* 2013.
- [62] L. Xu, M. Li, and A. R. Butt, "GERBIL: MPI + YARN," in *Cluster, Cloud and Grid Computing, IEEE/ACM Intl. Symposium on,* 2015.
- [63] N. Mishra, C.-C. Lin, and H.-T. Chang, "A cognitive oriented framework for IoT big-data management prospective," in *Communication Problem-Solving, IEEE Intl. Conf. on.* IEEE, 2014.
- [64] E.-E. Durham, A. Rosen, R. W. Harrison *et al.*, "A model architecture for big data applications using relational databases," in *Big Data, IEEE Intl. Conf. on,* 2014.
- [65] A. Chebotko, 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] A. Ochian, G. Suci, O. Fratu, and V. Suci, "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. Habib ur Rehman, C. S. 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. Kousiouris, 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, I. Paik, J. Zhang, T. Siriweera, and K. R. Koswatta, "Ontology-based workflow generation for intelligent big data analytics," in *Web Services, IEEE Intl. Conf. on,* 2015.
- [73] M. Westerlund, U. Hedlund, G. Pulkkis, and K.-M. Björk, "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, R. Kazman, and S. Haziye, "Strategic prototyping for developing big data systems," *IEEE Software,* 2016.
- [76] W. Zhang, L. Xu, Z. Li, Q. Lu, and Y. Liu, "A deep-intelligence framework for online video processing," *IEEE Software,* 2016.
- [77] M. M. Bersani, F. Marconi, D. A. Tamburri, P. Jamshidi, and A. Nodari, "Continuous architecting of stream-based systems," in *13th Working IEEE/IFIP Conf. on Software Architecture,* 2016.
- [78] A. Zimmermann, B. Gonen, R. Schmidt, E. El-Sheikh, S. Bagui, and N. Wilde, "Adaptable enterprise architectures for software evolution of SmartLife ecosystems," in *Intl. Enterprise Distributed Object Computing Conf. Workshops and Demonstrations.* IEEE, 2014.
- [79] B. Li, M. Grechanik, and D. Poshvyanyk, "Sanitizing and minimizing databases for software application test outsourcing," in *Intl. Conf. on Software Testing, Verification and Validation.* IEEE, 2014.
- [80] C. L. Wu, T. C. Chiang, L. C. Fu, 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] M. Thangaraj and S. Anuradha, "State of art in testing for big data," in *IEEE Intl. Conf. on Computational Intelligence and Computing Research,* 2015.
- [82] M. O. Gökalp, A. Koçyigit, and P. E. Eren, "A cloud based architecture for distributed real time processing of continuous queries," in *Euromicro Conf. on Software Engineering and Advanced Applications,* 2015.
- [83] D. Wu, L. Zhu, X. Xu, S. Sakr, D. Sun, and Q. Lu, "Building pipelines for heterogeneous execution environments for big data processing," *IEEE Software,* 2016.
- [84] D. G. Tesfagiorgish and L. JunYi, "Big data transformation testing based on data reverse engineering," in *Intl. Conf. on Ubiquitous Intelligence and Computing, Intl.*

- Conf. on Autonomic and Trusted Computing, Intl. Conf. on Scalable Computing and Communications (UIC-ATC-ScalCom)*, 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, J. G. Lou, H. Zhang, H. Lin, H. Lin, and T. Qin, "An empirical study on quality issues of production big data platform," in *IEEE/ACM Intl. Conf. on Software Engineering*, 2015.
- [88] A. Samuel, M. I. Sarfraz, H. Haseeb, S. Basalamah, and A. Ghafoor, "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, I. Zavorin, Z. Lim, S. Muthiah, L. Zhao, C. T. Lu, P. Butler, R. P. Khandpur, Y. Fayed, and N. Ramakrishnan, "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, W. Yu, Y. Hu, K. Wang, J. Wang, and S. Li, "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. Haziyevev, "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] N. Wilder, J. M. Smith, and A. Mockus, "Exploring a framework for identity and attribute linking across heterogeneous data systems," in *Proceedings of the 2nd Intl. Workshop on BIG Data Software Engineering*. ACM, 2016.
- [95] J. Klein, I. Gorton, L. Alhmoud, J. Gao, C. Gemici, R. Kapoor, P. Nair, and V. Saravagi, "Model-driven observability for big data storage," in *13th Working IEEE/IFIP Conf. on Software Architecture*, 2016.
- [96] G. Gousios, D. Safaric, and J. Visser, "Streaming software analytics," in *Proceedings of the 2nd Intl. Workshop on BIG Data Software Engineering*. ACM, 2016.
- [97] M. Guerriero, S. Tajfar, D. A. Tamburri, and E. Di Nitto, "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, D. Arruda, and N. H. Madhavji, "Understanding quality requirements in the context of big data systems," in *Proceedings of the 2nd Intl. Workshop on BIG Data Software Engineering*. ACM, 2016.
- [99] S. Marchal, X. Jiang, R. State, and T. Engel, "A big data architecture for large scale security monitoring," in *IEEE Intl. Congress on Big Data*, June 2014, pp. 56–63.
- [100] F. Zhang, J. Cao, S. U. Khan, K. Li, 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, H. Wang, Y. Lu, T. Li, Y. Guang, C. Liu, E. Edrosa, H. Li, and N. Rishe, "TerraFly Geocloud: An online spatial data analysis and visualization system," *ACM Transactions on Intelligent System and Technology*, 2015.
- [102] Z. Xu, X. Wei, X. Luo, Y. Liu, L. Mei, C. Hu, and L. Chen, "Knowle: A semantic link network based system for organizing large scale online news events," *Future Generation Computer Systems*, 2015.
- [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, P.-F. Li, L.-L. Tian, Y.-M. Qian, and J.-S. Li, "Design and Development of a Medical Big Data Processing System based on Hadoop," *Journal of medical systems*, 2015.
- [105] M. A. Saleem, Y.-K. Lee, and S. Lee, "Trajectory patterns mining towards lifecare provisioning," *Wireless Personal Communications*, 2014.
- [106] S. J. Rysavy, D. Bromley, and V. Daggett, "DIVE: A graph-based visual-analytics framework for big data," *IEEE Computer Graphics and Applications*, 2014.
- [107] A. Naseer, B. Y. Alkazemi, and E. U. 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, M. Qiu, 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] J. Ding, 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. Šendelj, I. Ognjanović, E. Ammenwerth, and W. Hackl, "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. Forkan, I. Khalil, A. Ibaida, and Z. Tari, "BDCaM: Big data for context-aware monitoring - A Personalized Knowledge Discovery Framework for Assisted Healthcare," *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. Müller, L. Bernard, and D. Kadner, "Moving code-sharing geoprocessing logic on the web," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013.
- [115] T. Shelton, A. Poorthuis, M. Graham, and M. Zook, "Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of big data," *Geoforum*, 2014.
- [116] G. Cao, S. Wang, M. Hwang, 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, I. Allison, and H. González-Vélez, “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 and L. Di, “Building open environments to meet big data challenges in Earth sciences,” in *Big Data : Techniques and Technologies in Geoinformatics*. CRC Press, 2013.
- [120] S. Fang, L. Da Xu, Y. Zhu, J. Ahati, H. Pei, J. Yan, and Z. Liu, “An integrated system for regional environmental monitoring and management based on internet of things,” *Industrial Informatics, IEEE Transactions on*, 2014.
- [121] C. Ledur, D. Griebler, 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, K. M. Anderson, M. Kogan, and L. Palen, “EPIC-OSM: A software framework for OpenStreetMap Data Analytics,” in *Hawaii Intl. Conf. on System Sciences*, 2016.
- [123] N. Pelekis, Y. Theodoridis, and D. Janssens, “On the management and analysis of our lifestyles,” *SIGKDD Explor. Newsl.*, 2014.
- [124] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, “CAP: Community Activity Prediction based on big data analysis,” *Network, IEEE*, 2014.
- [125] S. D’Oca and T. Hong, “Occupancy schedules learning process through a data mining framework,” *Energy and Buildings*, 2015.
- [126] I. Widjaja, P. Russo, C. Pettit, 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] W. Q. Wang, X. Zhang, J. 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, L. N. Dunn, M. D. Sohn, A. Mercado, C. Custodio, 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] F. J. Wu, 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] W. Shi, Y. Zhu, J. Zhang, X. Tao, G. Sheng, Y. Lian, G. Wang, and Y. Chen, “Improving power grid monitoring data quality: An efficient machine learning framework for missing data prediction,” 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.
- [133] B. Cheng, S. Longo, F. Cirillo, M. Bauer, and E. Kovacs, “Building a big data platform for smart cities: Experience and lessons from Santander,” in *IEEE Intl. Congress on Big Data*, June 2015, pp. 592–599.
- [134] H. Li, D. Parikh, Q. He, B. Qian, Z. Li, D. Fang, and A. Hampapur, “Improving rail network velocity: A machine learning approach to predictive maintenance,” *Transportation Research Part C: Emerging Technologies*, 2014.
- [135] A. Thaduri, D. Galar, 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. Kwoczek, S. D. Martino, T. Rustemeyer, and W. Nejdl, “An architecture to process massive vehicular traffic data,” in *Intl. Conf. on P2P, Parallel, Grid, Cloud and Internet Computing*, 2015.
- [138] R. O. Sinnott, L. Morandini, and S. Wu, “SMASH: A cloud-based architecture for big data processing and visualization of traffic data,” in *IEEE Intl. Conf. on Data Science and Data Intensive Systems*, 2015.
- [139] J. Yang and J. Ma, “A big-data processing framework for uncertainties in transportation data,” in *Fuzzy Systems, IEEE Intl. Conf. on*, Aug 2015, pp. 1–6.
- [140] M. Yesudas, G. Menon, and V. Ramamurthy, “Intelligent operational dashboards for smarter commerce using big data,” *IBM Journal of Research and Development*, 2014.
- [141] L. Shi, F. Lin, T. Yang, J. Qi, W. Ma, 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, J. Gao, and C. Vuppalapati, “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. Suci, C. Dobre, V. Suci, G. Todoran, A. Vulpe, 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] K. Tao, C. Hauff, G. J. Houben, F. Abel, and G. Wachsmuth, “Facilitating Twitter data analytics: Platform, language and functionality,” in *Big Data, Intl. Conf. on*. IEEE, 2014.
- [147] A. Immonen, P. Pääkkönen, and E. Ovaska, “Evaluating the quality of social media data in big data architecture,” *IEEE Access*, 2015.
- [148] I. D. Addo, D. Do, R. Ge, and S. I. Ahamed, “A reference architecture for social media intelligence applications in the cloud,” in *Computer Software and*

- Applications Conference.* IEEE, 2015.
- [149] E. W. Patton, P. Seyed, P. Wang, L. Fu, F. J. Dein, R. S. Bristol, and D. L. McGuinness, "SemantEco: A semantically powered modular architecture for integrating distributed environmental and ecological data," *Future Generation Computer Systems*, 2014.
- [150] C. Beal and J. Flynn, "Toward the digital water age: Survey and case studies of Australian water utility smart-metering programs," *Utilities Policy*, 2015.
- [151] E. Moguel, J. C. Preciado, F. Sanchez-Figueroa, M. Preciado, J. Hernandez *et al.*, "Multilayer big data architecture for remote sensing in eolic parks," *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015.
- [152] D. Dutta and I. Bose, "Managing a big data project: The case of Ramco Cements limited," *Intl. Journal of Production Economics*, 2015.
- [153] N. Kushiro, S. Matsuda, and K. Takahara, "Model oriented system design on big-data," *Procedia Computer Science*, 2014.
- [154] J. Lee, H.-A. Kao, and S. Yang, "Service innovation and smart analytics for industry 4.0 and big data environment," *Procedia CIRP*, 2014.
- [155] C. Li, Y. Liu, R. Li, and H. Zhang, "Research and application of one-key publishing technologies for meteorological service products," in *IEEE Intl. Conf. on Big Data Analysis*, 2016.
- [156] M. J. Divn, Y. B. Saibene, M. D. L. . Martn, M. L. Belmonte, G. Lafuente, and 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.
- [157] L. Zhang, "A framework to model big data driven complex cyber physical control systems," in *Automation and Computing, Intl. Conf. on.* IEEE, 2014.
- [158] C. Cecchinell, S. Mosser, and P. Collet, "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.
- [159] L. Zhang, "Designing big data driven cyber physical systems based on AADL," in *Systems, Man and Cybernetics, IEEE Intl. Conf. on.*, 2014.
- [160] M. Rahmes, G. Lemieux, K. Fox, and C. Casseus, "Multi-disciplinary ontological geo-analytical incident modeling," in *IEEE Consumer Communications and Networking Conference*, 2015.
- [161] J. Klein, R. Buglak, D. Blockow, T. Wuttke, and B. Cooper, "A reference architecture for big data systems in the National Security domain," in *Proceedings of the 2nd Intl. Workshop on Big Data Software Engineering*. ACM, 2016.
- [162] W. Akio Goya, M. Risse de Andrade, A. Carvalho Zucchini, N. Mimura Gonzalez, R. de Fatima Pereira, K. Langona, T. C. Melo de Brito Carvalho, J.-E. Mangs, 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.
- [163] R. Dutta, A. Morshed, J. Aryal, C. D'este, and A. Das, "Development of an intelligent environmental knowledge system for sustainable agricultural decision support," *Environmental Modelling & Software*, 2014.
- [164] N. Sun, J. Morris, J. Xu, X. Zhu, and M. Xie, "iCARE: A framework for big data-based banking customer analytics," *IBM Journal of Research and Development*, 2014.
- [165] C. Restrepo-Arango, A. Henao-Chaparro, and C. Jiménez-Guarín, "Using the web to monitor a customized unified financial portfolio," in *Advances in Conceptual Modeling*. Springer, 2012.
- [166] D. Ning, P. Chen, G. Yuan, J. Xu, and L. Xu, "Research on Warship Communication Operation and Maintenance Management Based on Big Data," in *Cloud Computing and Big Data, Intl. Conf. on.* IEEE, 2014.
- [167] E. Sciacca, C. Pistagna, U. Becciani, A. Costa, P. Massimino, S. Riggi, F. Vitello, M. Bandieramonte, and M. Krokos, "Towards a big data exploration framework for astronomical archives," in *High Performance Computing & Simulation, Intl. Conf. on.* IEEE, 2014.
- [168] 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.
- [169] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, 2014.
- [170] C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information Sciences*, 2014.
- [171] J. S. Ward and A. Barker, "Undefined by data: A survey of big data definitions," *arXiv preprint arXiv:1309.5821*, 2013.
- [172] I. Gorton, A. B. Bener, and A. Mockus, "Software Engineering for Big Data Systems," *IEEE Software*, 2016.
- [173] 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.
- [174] C. E. Cuesta, M. A. Martínez-Prieto, and J. D. Fernández, "Towards an architecture for managing big semantic data in real-time," in *Software Architecture*. Springer, 2013.
- [175] R. S. Pressman, *Software Engineering: A Practitioner's Approach*. Palgrave Macmillan, 2005.
- [176] 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.