

# Conference Paper Title\*

\*Note: Sub-titles are not captured in Xplore and should not be used

1<sup>st</sup> Given Name Surname  
*dept. name of organization (of Aff.)*  
*name of organization (of Aff.)*  
City, Country  
email address or ORCID

2<sup>nd</sup> Given Name Surname  
*dept. name of organization (of Aff.)*  
*name of organization (of Aff.)*  
City, Country  
email address or ORCID

3<sup>rd</sup> Given Name Surname  
*dept. name of organization (of Aff.)*  
*name of organization (of Aff.)*  
City, Country  
email address or ORCID

4<sup>th</sup> Given Name Surname  
*dept. name of organization (of Aff.)*  
*name of organization (of Aff.)*  
City, Country  
email address or ORCID

5<sup>th</sup> Given Name Surname  
*dept. name of organization (of Aff.)*  
*name of organization (of Aff.)*  
City, Country  
email address or ORCID

6<sup>th</sup> Given Name Surname  
*dept. name of organization (of Aff.)*  
*name of organization (of Aff.)*  
City, Country  
email address or ORCID

**Abstract**—This document is a model and instructions for L<sup>A</sup>T<sub>E</sub>X. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. **\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

**Index Terms**—component, formatting, style, styling, insert

## I. REQUIREMENTS SPECIFICATION

Precursor to theorizing about the potential of microservices patterns for big data systems, we need to define what we mean by big data systems and what are the requirements of these systems. System and software requirements come in different flavour and can range from a sketch on a napkin to formal (mathematical) specifications. Therefore, we first need to identify what kind of requirements is the most suitable for the purposes of this study. To answer this question, we first explored the body of evidence to understand the current classification of software requirements.

There's been various attempts to defining and classifying software and systems requirements. For instance, Sommerville ([1]) classified requirements into three levels of abstraction that are namely 1) user requirements, 2) system requirements and 3) design specifications. The author then mapped these requirements against user acceptance testing, integration testing and unit testing. While this could satisfy the requirements of this study, we opted for a more general framework provided by Laplante ([2]). In Laplante's approach, requirements are categorized into three categories of 1) functional requirements, 2) non-functional requirements, and 3) domain requirements.

Our objective is to define the high-level requirements of big data systems, thus we do not seek to explore 'non-functional' requirements. Non-functional requirements are emerged from the particularities of an environment, such as a banking sector and do not correlate to our study. Therefore, the type of

requirements we are looking for is functional and domain requirements.

After clarifying the type of requirements, we then explored the body of evidence to realize the general requirements of big data systems. Indeed, the most discussed characteristics of big data systems are the popular 5Vs which are velocity, veracity, volume, Variety and Value ([3], [4], [5], [6], [7], [8]). Many researchers such as Nadal et al. ([9]) have underpinned their artifact development on these characteristics and requirements that emerge from them.

In an extensive effort, NIST Big Data Public Working Group embarked on a large scale study to extract requirements from variety of application domains such as Healthcare and Life Sciences, Commercial, Energy, Government, and Defense. The result of this study was the formation of general requirements under seven categories. In another effort by Volk et al. ([10]), 9 use cases for big data projects are identified by collecting theories and use cases from the literature and categorizing them using a hierarchical clustering algorithm. Bashari et al. ([11]) focused on the security and privacy requirements of big data systems, Yu et al. presented the modern components of big data systems [12], Eridaputra et al. ([13]) created a generic model for big data requirements using goal oriented approaches, and Al-jaroodi et al. ([14]) investigated general requirements to support big data software development.

We've also studied the reference architectures developed for big data systems to understand general requirements. In one study, Ataei et al. ([15]) assessed the body of evidence and presented with a comprehensive list of big data reference architectures. This study helped us realized the spectrum of big data reference architectures, how they are designed and the general set of requirements.

By analyzing these studies and by evaluating the design and requirement engineering required for big data reference architectures, we created a set of high-level requirements based on big data characteristics. We have then looked for

Identify applicable funding agency here. If none, delete this.

a rigorous approach to present these requirements. There are numerous approaches used for requirement representation including informal, semiformal and formal methods. For the purposes of this study, we opted for an informal method because it's a well established method in the industry and academia ([16]).

Our approach follows the guidelines explained in ISO/IEC/IEEE standard 29148 for representing functional requirements. Our requirement representation is organized in system modes, that is we explain the major components of the system and then describe the requirements. This approach is inspired by the requirement specification expressed for NASA WIRE (wide-field infrared explorer) system explained in [2]. We also taken inspiration from Software Engineering Body of Knowledge Version ([17]).

Taking all into consideration, we categorized our requirements based on the major characteristics of big data, that is value, variety, velocity, veracity, and volume ([?]), plus . These requirements are as followings:

## REFERENCES

- [1] I. Sommerville, *Software Engineering, 9/E*. Pearson Education India, 2011.
- [2] P. A. Laplante, *Requirements engineering for software and systems*. Auerbach Publications, 2017.
- [3] Y. Demchenko, C. De Laat, and P. Membrey, "Defining architecture components of the big data ecosystem," in *2014 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2014, Conference Proceedings, pp. 104–112.
- [4] J. Bughin, "Big data, big bang?" *Journal of Big Data*, vol. 3, no. 1, p. 2, 2016.
- [5] M. Bahrami and M. Singhal, *The role of cloud computing architecture in big data*. Springer, 2015, pp. 275–295.
- [6] B. B. Rad and P. Ataei, "The big data ecosystem and its environs," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 17, no. 3, p. 38, 2017.
- [7] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co., 2015.
- [8] H.-M. Chen, R. Kazman, and S. Haziye, "Agile big data analytics development: An architecture-centric approach," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016, Conference Proceedings, pp. 5378–5387.
- [9] S. Nadal, V. Herrero, O. Romero, A. Abelló, X. Franch, S. Vansumneren, and D. Valerio, "A software reference architecture for semantic-aware big data systems," *Information and software technology*, vol. 90, pp. 75–92, 2017.
- [10] M. Volk, D. Staegemann, I. Trifonova, S. Bosse, and K. Turowski, "Identifying similarities of big data projects—a use case driven approach," *IEEE Access*, vol. 8, pp. 186 599–186 619, 2020.
- [11] B. Bashari Rad, N. Akbarzadeh, P. Ataei, and Y. Khakbiz, "Security and privacy challenges in big data era," *International Journal of Control Theory and Applications*, vol. 9, no. 43, pp. 437–448, 2016.
- [12] J.-H. Yu and Z.-M. Zhou, "Components and development in big data system: A survey," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 51–72, 2019.
- [13] H. Eridaputra, B. Hendradjaya, and W. D. Sunindyo, "Modeling the requirements for big data application using goal oriented approach," in *2014 international conference on data and software engineering (ICODSE)*. IEEE, 2014, pp. 1–6.
- [14] J. Al-Jaroodi and N. Mohamed, "Characteristics and requirements of big data analytics applications," in *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 2016, pp. 426–432.
- [15] P. Ataei and A. T. Litchfield, "Big data reference architectures, a systematic literature review," 2020.
- [16] M. Kassab, C. Neill, and P. Laplante, "State of practice in requirements engineering: contemporary data," *Innovations in Systems and Software Engineering*, vol. 10, no. 4, pp. 235–241, 2014.
- [17] A. Abran, J. W. Moore, P. Bourque, R. Dupuis, and L. Tripp, "Software engineering body of knowledge," *IEEE Computer Society, Angela Burgess*, p. 25, 2004.

Volume	<ul style="list-style-type: none"> <li>1) The system needs to support asynchronous, streaming, and batch processing to collect data from centralized, distributed, and cloud data sources, and sensors, instrument and other IOT devices</li> <li>2) The systems needs to be able to process large heterogenous data with varying schemas</li> <li>3) The system needs to provide a scalable storage for massive data sets</li> </ul>
Velocity	<ul style="list-style-type: none"> <li>1) The system needs to support slow, bursty, and high-throughput data transmission between data sources and computing clusters</li> <li>2) The systems needs to stream data to data consumers in a timely manner</li> <li>3) The system needs to able to ingest multiple, continuous, time varying data streams</li> <li>4) The system shall support fast search from streaming and processed data with high accuracy and relevancy</li> <li>5) The system should be able to process data in real-time or near real-time manner</li> </ul>
Variety	<ul style="list-style-type: none"> <li>1) The system needs to support data in various formats ranging from structured to semi-structured and unstructured graph, web, text, document, timed, spatial, multimedia, simulation, instrumental, and geo-spatial data.</li> <li>2) The system needs to support aggregation, standardization, and normalization of data from disparate sources</li> </ul>
Value	<ul style="list-style-type: none"> <li>1) Needs to support diversified compute-intensive, statistical and graph analytic processing, and machine learning techniques.</li> <li>2) Needs to support batch and real-time analytic processing.</li> <li>3) Needs to support legacy and advanced executable programming: applications, tools, utilities, and libraries (software).</li> <li>6) Needs to support diversified output file formats for visualization, rendering, and reporting.</li> <li>7) Needs to support visual layout for results presentation. 8)Needs to support rich user interface for access using browser, visualization tools.</li> <li>9) Needs to support high-resolution, multi-dimension layer of data visualization.</li> <li>10) Needs to support high-resolution, multi-dimension layer of data visualization.</li> <li>11) Needs to support streaming results to clients.</li> <li>12) The BDA shall be capable of supporting descriptive analytics.</li> <li>13) The BDA shall be capable of supporting predictive and prescriptive analytics.</li> </ul>
Security & Privacy	<ul style="list-style-type: none"> <li>1) Needs to protect and preserve security and privacy of sensitive data. 2) Needs to support sandbox, access control, and multi-level, policy-driven authentication on protected data.</li> </ul>
Veracity	<ul style="list-style-type: none"> <li>1) Needs to support data quality curation including pre-processing, data clustering,classification, reduction, and format transformation. 2) Needs to support data life cycle and long-term preservation policy, including data provenance. 3) Needs to support data validation. 4) Needs to support human annotation for data validation. 5) Needs to support persistent identifier and data traceability. 7) The BDA shall provide mechanisms to measure data quality. 8) The BDA shall provide mechanisms for tracing data liveliness. 10) Needs to support prevention of data loss or corruption.</li> </ul>
Variability	<ul style="list-style-type: none"> <li>1) The BDA shall provide adaptation mechanisms to schema evolution. 2) The BDA shall provide mechanisms for automatic inclusion of new data sources.</li> </ul>