

12-15-2019

Unpacking the Evolution of Big Data and Organizing: An Illustrative Case Study

B. Veeresh Thummadi
University of Limerick, veeresh.thummadi@lero.ie

Follow this and additional works at: <https://aisel.aisnet.org/digit2019>

Recommended Citation

Thummadi, B. Veeresh, "Unpacking the Evolution of Big Data and Organizing: An Illustrative Case Study" (2019). *DIGIT 2019 Proceedings*. 1.
<https://aisel.aisnet.org/digit2019/1>

This material is brought to you by the Diffusion Interest Group In Information Technology at AIS Electronic Library (AISeL). It has been accepted for inclusion in DIGIT 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Unpacking the Evolution of Big Data and Organizing: An Illustrative Case Study

Research Idea

B. Veeresh Thummadi

Lero the Irish Software Research Centre,
University of Limerick, Limerick, Ireland.
veeresh.thummadi@lero.ie

Abstract

In the last decade we have witnessed the growth of emerging technologies like big data that have challenged our understanding on how we carry out work and processes in organizations. I have collected over 10 years of data containing over 1.6 million comments from a popular big data community to understand how organizations resolve the tensions between the moving parts of the firm and big data technologies. As big data technologies are complex, organizations can struggle in terms of transitioning to big data systems unlike the traditional IT systems, hence I plan on conducting a qualitative study by analyzing the comments to understand the evolution process of big data technologies. I plan on using the routine lens to study this phenomenon as it can shed light into the mechanisms through which organizations resolve tensions while implementing big data systems. Specifically, the proposed research study has two goals. First, I wish to develop a formal process theory about the adoption of big data innovations and organizing using grounded theory approach. Second, through this study I would like to unpack how organizations resolve tensions between while implementing big data technologies.

Keywords: bigdata, open source, case study, routines, evolution.

Unpacking the Evolution of Big Data and Organizing: An Illustrative Case Study

Research Idea

The concept of big data comes from early 2000's where the word has been extensively used to refer to large volumes amounts of data. Research on big data spans across multiple disciplines from natural sciences to social sciences and takes place under different headings such as data science and data analytics (Jin, Wah, Cheng, & Wang, 2015). In the last decade, big data has received considerable attention from industry and academic scholars and has led to new research centres across the world. A substantial part of this big data research paradigm focuses on data processing, data extraction, pattern recognition, industrial applications, firm performance and economic development (Provost & Fawcett, 2013; Tambe, 2014). However, very little attention has been paid on the dynamics through which the big data platforms evolve and their impact on the organizational day-to-day operations. Hence I ask: How does big data evolve and affect the work processes in organizations? More specifically, what are the mechanisms through which actors grapple the existing work processes and resolve the tensions between big data growth and organizing?

Despite wide scale popularity, big data revolution comes up with many challenges that are usually not covered by the popular media such as complexity of the processing files, indexing and security issues in the development of big data platforms. Here organizations using big data platforms would face two types of challenges: 1) how to replace the existing work processes with new features of the technology? And 2) how to successfully minimize the frictions between moving parts of the organization? Management scholars suggest use the organizational routines lens to understand how advanced technologies like big data are successfully adapted in the organizations i.e., routines can be used to unpack the mechanisms by which organizations reach stability and change (D'Adderio, 2014; Feldman & Pentland, 2003; Pink, Sumartojo, Lupton, & Heyes La Bond, 2017). Using routine lens would be fruitful in this context as organizations are still struggling in adopting big data technologies unlike other Information Technology (IT) systems such as ERP or SAP (to name a few) (Berente, Lyytinen, Yoo, & King, 2016; D'Adderio, 2014; Turner & Rindova, 2012). My line of argument follows this logic and specifically looks at how organizations adapt and reorganize the existing work practises. To address these questions, I have collected data from a popular big data community, Hadoop Distributed File System (HDFS). HDFS was originally built as infrastructure for the Apache Nutch pilot project, envisioned by Doug Cutting for creating a search engine like Google (Tambe, 2014). HDFS is also one of the Apache hadoop core projects and has a master/slave architecture setup for carrying out its operations and was designed to be highly fault-tolerant. For example, when working with large datasets, files are broken down into small chunks and the operations are performed at multiple computers (also called nodes) and HDFS allows operations to run successfully despite failure at any of the nodes (see Figure 1).

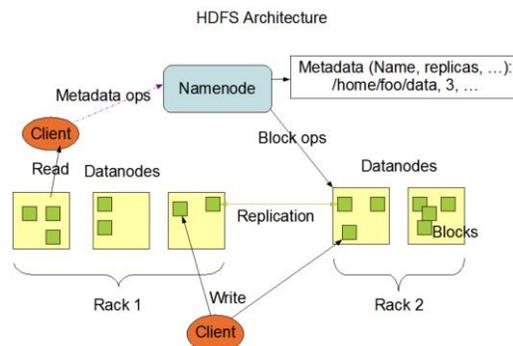


Figure 1: Snapshot of HDFS architecture (Courtesy: www.apache.org)

The HDFS project allows thousands of hadoop users and organizations to share hadoop-related problems, contribute their knowledge on the issues, and to learn from other people's understanding/comments. In addition, a project lead is assigned to the HDFS project who is responsible for managing the project, assigning issues, checking the resolutions provided by the assignee and marking them as fixed. Thus, the HDFS project can provide insights into the developer-organization interactions and the challenges organizations face during implantation. Overall I collected data on 12,230 issues and 1.6 million comments over 10 years by multiple participants/users on this project. Further, I plan on conducting interviews with various stakeholders to better understand how HDFS systems evolved in time and the challenges faced while transitioning to HDFS. The main focus of this study is to look at how data storage and processing facilities evolved over time. I plan on analysing the data by exploring the archival data to understand the challenges in the product design and organizational work units and their relationships. First, I would like to systematically understand how HDFS has evolved in time. This can provide some insight on how the big data product features are shaped by the organizations and their internal processes. Then, I plan on conducting the data analysis in an iterative fashion travelling back and forth between the data, constructs and literature (Strauss & Corbin, 1994). Using grounded theory approach I will analyze the qualitative data in a systematic manner. First few rounds of analysis will be done for familiarization stage, which will be later revisited. Given my focus in understanding the stability and change in routine in big data technologies I will perform first order analysis for discovering themes and patterns in the events described by the actors.

Specifically this research aims to provide two main contributions. First, through this research I would like to advance the state-of-art of the understanding on how big data platforms evolve. As big data platforms are complex the understanding of the evolution process can provide an understanding of how emerging technologies can be developed using distributed and open source development models. Second, by unpacking the relationship between work process and big data growth, the research aims to tease out the mechanisms by which organizations resolve the tensions between the moving parts of the organization. Previous research on organizational routines suggests that artifacts can be crucial in resolving the tensions (Berente, Lyytinen, Yoo, & King, 2016; D'Adderio, 2014; Turner & Rindova, 2012), however, we are yet to see how organizations cope with a radically innovative technology like big data. This can be useful for organizations for reducing the friction between different organizational units while implementing big data technologies.

Acknowledgements

This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094.

References

- Berente, N., Lyytinen, K., Yoo, Y., & King, J. L. 2016. Routines as Shock Absorbers During Organizational Transformation: Integration, Control, and NASA's Enterprise Information System. *Organization Science*.
- D'Adderio, L. 2014. The replication dilemma unravelled: how organizations enact multiple goals in routine transfer. *Organization Science*, 25(5): 1325-1350.
- Feldman, M. S., & Pentland, B. T. 2003. Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly*, 48(1): 94-121.
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. 2015. Significance and challenges of big data research. *Big Data Research*, 2(2): 59-64.
- Pink, S., Sumartojo, S., Lupton, D., & Heyes La Bond, C. 2017. Mundane data: The routines, contingencies and accomplishments of digital living. *Big Data & Society*, 4(1): 2053951717700924.
- Provost, F., & Fawcett, T. 2013. Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1): 51-59.
- Strauss, A., & Corbin, J. 1994. Grounded theory methodology. *Handbook of qualitative research*: 273-285.
- Tambe, P. 2014. Big data investment, skills, and firm value. *Management Science*, 60(6): 1452-1469.
- Turner, S. F., & Rindova, V. 2012. A balancing act: How organizations pursue consistency in routine functioning in the face of ongoing change. *Organization Science*, 23(1): 24-46.