

## Highlights

### **Terramycelium: A Domain-driven Reference Architecture for Big Data Systems**

Author One, Author Two, Author Three

- Research highlight 1
- Research highlight 2

# Terramycelium: A Domain-driven Reference Architecture for Big Data Systems

Author One<sup>a</sup>, Author Two<sup>b</sup>, Author Three<sup>a,b</sup>

<sup>a</sup>*Department One, Address One, City One, 00000, State One, Country One*

<sup>b</sup>*Department Two, Address Two, City Two, 22222, State Two, Country Two*

---

## Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

*Keywords:* keyword one, keyword two

*PACS:* 0000, 1111

*2000 MSC:* 0000, 1111

---

## 1. Introduction

The advent of the internet and widespread use of digital devices have sparked a profound shift in connectivity and data creation, leading to an era marked by a rapid growth in data. This period is characterised by the extensive expansion of data, which presents difficulties for traditional data processing systems and necessitates inventive methods in data architecture [1, 2]. The vast amount, variety, and rapid generation of data in the current digital environment necessitate innovative solutions, particularly in the field of Big Data (BD).

Data needs have dramatically evolved, transitioning from basic business intelligence (BI) functions, like generating reports for risk management and compliance, to incorporating machine learning across various organisational facets [3]. These range from product design with automated assistants to personalised customer service and optimised operations. Also, as machine

learning becomes more popular, application development needs to change from rule-based, deterministic models to more flexible, probabilistic models that can handle a wider range of outcomes and need to be improved all the time with access to the newest data. This evolution underscores the need to reevaluate and simplify our data management strategies to address the growing and diverse expectations placed on data.

Currently, the success rate of BD projects is low. Recent surveys have identified the fact that current approaches to big data do not seem to be effectively addressing these expectations. According to a survey conducted by [4], only 13% of organisations are highly successful in their data strategy. Additionally, a report by NewVantage Partners reveals that only 24% of organisations have successfully converted to being data-driven, and a measly 30% have a well-established big data strategy. These observations, additionally corroborated by research conducted by McKinsey & Company (analytics2016age) and Gartner (Nash), emphasise the difficulties of successfully using big data in the industry. These difficulties include the lack of a clear understanding of how to extract value from data, the challenge of integrating data from multiple sources, data architecture, and the need for skilled data analysts and scientists.

Without a well-established big data strategy, companies may struggle to navigate these challenges and fully leverage the potential of their data. One effective artefact to overcome some of these challenges is Reference Architectures (RAs) [5]. RAs extract the essence of the practice as a series of patterns and architectural constructs and manifest it through high-level semantics. This allows stakeholders to refrain from reinventing the wheel and instead focus on utilising existing knowledge and best practices to harness the full potential of their data. While there are various BD RAs available to help practitioners design their BD systems, these RAs are overly centralised, lack attention to cross-cutting concerns such as privacy, security, and metadata, and may not effectively handle the proliferation of data sources and consumers.

To this end, this study presents TerrMycelium, a distributed RA designed specifically for BD systems with a focus on domain-driven design. TerrMycelium seeks to surpass the constraints of current RAs by utilising domain-driven and distributed approaches derived from contemporary software engineering. This method aims to improve the ability of BD systems to scale, be maintained, and evolve, surpassing the constraints of traditional monolithic data architectures.

The paper is structured as follows: Section 2 provides an overview of the foundational concepts and technologies pertinent to BD reference architecture, aiming to forge a conceptual framework that is required for this paper. An overview of the existing research on the topic is presented in Section 3. The significance of reference architectures in the context of big data is explored in Section 4. Section 5 details the software and system requirements necessary for implementing the proposed architecture. Section 6 delves into the theoretical foundation underpinning the challenges in contemporary big data systems. The design and development of the TerrMycelium artifact are described in Section 7. Section 8 examines the evaluation findings, their implications, limitations, and relevance to existing and future research. Finally, Section 9 summarizes the main contributions of the study, its practical implications, and suggests directions for future research.

## **2. Background**

## **3. Related Work**

## **4. Why Reference Architectures**

## **5. Software and System Requirements**

## **6. Theory**

## **7. Artifact**

## **8. Discussion**

## **9. Conclusion**

## **References**

- [1] P. Ataei, A. Litchfield, Big data reference architectures: A systematic literature review, in: 2020 31st Australasian Conference on Information Systems (ACIS), IEEE, 2020, pp. 1–11. doi:10.5130/acis2020.bf.
- [2] B. B. Rad, P. Ataei, The big data ecosystem and its environs, International Journal of Computer Science and Network Security (IJCSNS) 17 (3) (2017) 38.
- [3] P. Ataei, A. Litchfield, Towards a domain-driven distributed reference architecture for big data systems, in: AMCIS 2023, 2023.

- [4] M. technology review insights in partnership with Databricks, Building a high-performance data organization (2021).  
URL <https://databricks.com/p/whitepaper/mit-technology-review-insights-report>
- [5] R. Cloutier, G. Muller, D. Verma, R. Nilchiani, E. Hole, M. Bone, The concept of reference architectures, Systems Engineering 13 (1) (2010) 14–27. doi:10.2514/6.2017-5118.