Abstract

Context: Big Data (BD) is a term that describes the large volume of data in various forms from different sources. Gleaning insights from this data can reveal transformative patterns, but the unprecedented scale has pushed traditional approaches to their limits. The growth of data is outpacing advances in data analytics, with an estimated 75% of BD projects failing in the last decade, often due to challenges in system development and data architecture. **Objective:** This paper aims to address these challenges by introducing a novel BD reference architecture called Terramycelium. Terramycelium absorbs principles from complex adaptive systems, domain-driven design, distributed systems, and event-driven systems. **Method:** The reference architecture was developed following guidelines for creating empirically grounded reference architectures. It was evaluated using two methods: a case-mechanism experiment and expert opinion. **Results:** The case-mechanism experiment results demonstrate Terramycelium's capability to meet BD system requirements while being highly maintainable and scalable. **Conclusions:** Terramycelium offers a promising new approach for designing BD systems that can handle the volume, velocity, and variety of modern data ecosystems. By leveraging domain-driven and distributed architectures, it aims to improve the scalability, maintainability, and adaptability of BD systems beyond the limitations of traditional monolithic data architectures.

1