10.18686/frim.v2i4.4353

Conceptual Model of Data Ecosystem for Non-Tech Small Organizations

Han Sun

Resource Pro (Qingdao) Co., Ltd, Qingdao City, Shandong Province, 266000

Abstract: Using data effectively is crucial for success in the current business world. Small non-tech organizations struggle with this due to limited resources, different systems, and constantly changing technology. To address these challenges, this study proposes a conceptual framework for designing a flexible and reusable data ecosystem, tailored specifically for small non-tech organizations. It also outlines the feedback loops that drive continuous improvement and innovation within the data ecosystem. Key considerations include data source integration, data quality management, and analytics capabilities. The proposed framework offers small non-tech organizations a structured approach to managing data assets effectively, enabling them to make informed decisions, optimize processes, and drive innovation.

Keywords: Data Ecosystem; Small Non-tech Organizations; Data Life-circle

1. Introduction

In recent years, there has been a significant focus on business analytics and the development of data infrastructure to drive decision-making processes in non-tech organizations and enhance their overall performance. Researchers such as Mayhew et al [1] have underscored the importance of organizations assigning purpose to their data and leveraging it to derive insights that add value.

Some companies that were born digital, like Google and Amazon, are skilled at using big data to their advantage [4]. This is also a great opportunity for other non-tech organizations to gain a competitive edge. However, small non-tech organizations face difficulties such as limited budgets and difficulties integrating old and new systems. Integrating traditional business practices with modern information systems presents additional challenges, requiring a balance between agility and control in decision-making, resource optimization, and communication channels [2]. Given these challenges, there is a critical need for novel data system solutions that offer flexibility and reusability to support organizational digital transformation efforts.

In my study, the concept of a "data ecosystem" within organizations is introduced, with a specific focus on business analytics and data infrastructure development. Decision-makers, particularly from traditional organizations, can leverage data ecosystems to identify potential threats to their market position, seize opportunities to adapt to evolving trends, and understand shifts in customer needs [3]. Oxford Dictionaries (2017) defines an ecosystem as “A biological system composed of all the organisms found in a particular physical environment, interacting with it and each other.” This definition extends metaphorically to describe intricate systems beyond the natural world. Much like natural ecosystems, the elements within this data ecosystem interact synergistically to perform various functions such as data production, management, storage, organization, analysis, and sharing.

2. Literature Review

In recent years, research has focused extensively on data ecosystems and their application across diverse industries, with a notable emphasis on the manufacturing sector. Tobias Riasanow’s study in 2017 [5] provides valuable insights into the automotive ecosystem, utilizing the e^3-value method to emphasize the central role of mobility service platforms, emerging disruptive technology providers, and collaborations between original equipment manufacturers (OEMs) and mobile payment providers.

Recent studies have examined how ecosystems work, including roles, governance, and ownership. In 2020, Andreas Hein's research highlighted the impact of ownership, value creation, and complementor autonomy on digital ecosystems [6]. Similarly, Lusch and Nambisan's 2015 work explored the shift from goods-based to service-based value creation in ecosystems, emphasizing the importance of collaboration and co-creation in promoting innovation and sustainable growth [7]. Furthermore, Tsujimoto et al.'s 2017 study describes ecosystems as self-organized or managerially designed social networks involving management, technology, and innovation [8].

Collectively, these studies underscore the increasing importance of data ecosystems across industries, shedding light on their structure, dynamics, and implications for organizational strategy, innovation, and sustainable development. However, the existing literature has primarily concentrated on large-scale implementations and case studies within tech-savvy or manufacturing-oriented environments. Small non-tech organizations face unique challenges, including limited resources, technical expertise, and scalability issues, which necessitate tailored solutions that prioritize simplicity, cost-effectiveness, and adaptability.

3. Methodology: Conceptual Model Approach

This study uses a method called conceptual modeling to create a clear and complete picture of how data moves through organizations. Conceptual modeling is the process that aims to reveal, analyze, and describe relevant entities of the subject domain, relationships between them, constraints they must satisfy, and their behavior [9].

Conceptual modeling offers several benefits in the context of building data infrastructure. Firstly, it aids in identifying and visualizing key concepts, facilitating a better understanding of how uncertain elements can influence businesses [10]. Moreover, conceptual modeling promotes collaboration and communication among stakeholders by providing a structured framework for discussing data-related strategies and objectives.

By employing the conceptual model methodology, the objective is to develop a robust framework for constructing data infrastructure. This model serves as a foundational blueprint for further business architecture design, incorporating a design rationale for logical reasoning and a causal-loop diagram for simulation purposes.

4. Model Development

Model planning within the conceptual model framework comprises several key components. By incorporating the components into the conceptual model methodology, the goal is to develop a comprehensive framework that supports data-driven decision-making, organizational growth, and the effective utilization of data assets. This architecture encompasses the entire data ecosystem, providing a clear pathway from data acquisition to analysis and utilization.

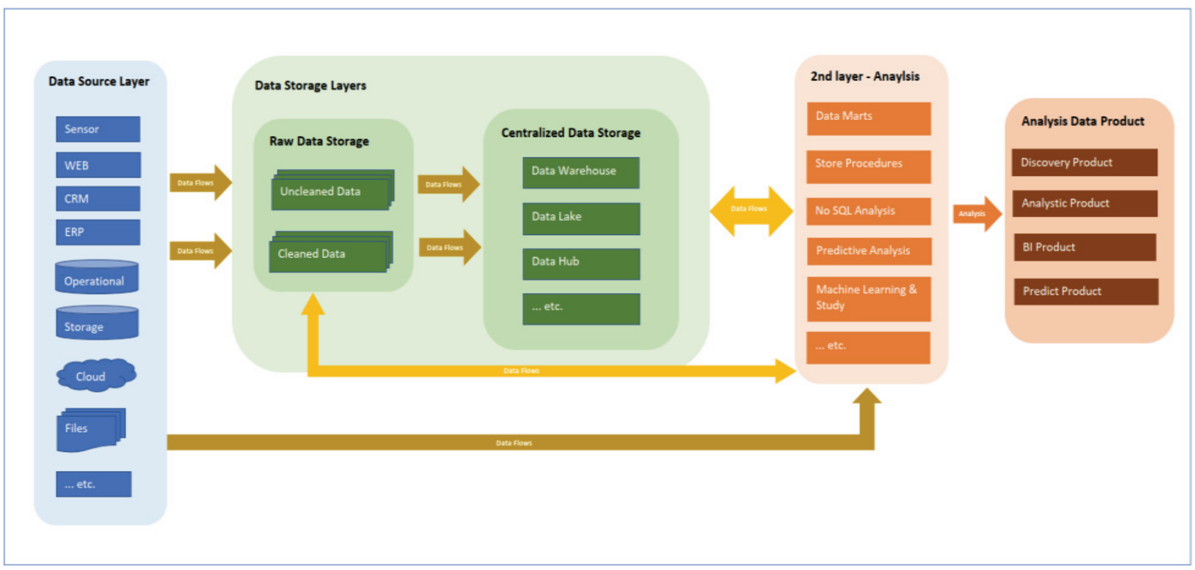

Figure 1: Overview of the Data Ecosystem Architecture

4.1 Data Source Layer

The data source layer is a foundational component of the data ecosystem, encompassing a diverse array of data sources. Data quality varies much from source to source even in the same domain, with significant differences in the accuracy and coverage of the data provided [11]. Despite these differences, the primary purpose of this layer is to collect as much data as possible, irrespective of quality. The goal is to gather all available data to create a comprehensive repository for the organization. However, it is important to note that this layer cannot encompass all possible data sources due to the unique focus areas, systems, and business needs of each organization. Therefore, the selection and integration of data sources must be tailored to align with the specific goals and operational requirements of the organization.

4.2 Data Flows

Data flows are important in the data ecosystem because they transfer data from its source to layers for processing, analysis, and storage. This makes sure that data moves smoothly and correctly through the system, maintaining its usefulness. Throughout this process, data may be changed to improve its quality, relevance, and consistency. During the data flow process, data can undergo various transformations to enhance its quality, relevance and consistence.

The data flows can be implemented across multiple platforms, including cloud-based and on-premises systems, to ensure seamless integration and movement of data throughout the ecosystem. This multi-platform approach enhances the system's flexibility and scalability, allowing it to adapt to changing organizational needs and technological advancements.

4.3 Data Storage Layers

In my architecture, the data storage is divided into two distinct layers to efficiently manage and organize the data before it is processed for analysis. However, organizations can adapt this model by combining these layers into one or further splitting them into additional layers, depending on their specific needs and requirements.

First Layer: Raw Data Storage

The first layer stores original data from different systems and sources, which may be messy or incomplete. At the same, it can work as a buffer and keeps the data in its original form for future use. One of the key functions of this layer is to perform initial preprocessing tasks. This includes basic data cleaning operations such as removing obvious errors, filtering out irrelevant information, and converting unstructured data into a more structured format. This preprocessing step ensures that the data is in a more manageable and consistent state before it moves to the next layer.

Second Layer: Centralized Data Storage

The second layer focuses on centralizing the data, making it readily available and accessible for analysis. This layer can encompass various forms of centralized storage solutions such as data warehouses, data lakes, and data hubs. By centralizing the data in this layer, organizations can ensure that all data, regardless of its source or format, is stored in a single, cohesive environment. This centralized storage not only improves data management and accessibility but also enhances data governance and security.

The distinction between the two storage layers helps in maintaining data integrity and facilitates efficient data processing workflows. The first layer acts as a repository for raw and transactional data, preserving its original form, while the second layer centralizes the processed data, making it suitable for analytical tasks. This structured approach to data storage enables organizations to effectively manage their data assets, supporting data-driven decision-making and strategic planning.

4.4 Data Analysis Layer

This layer is crucial because it transforms raw data into actionable insights, which can drive strategic decisions and operational improvements. The scope and depth of analysis within this layer can be tailored to fit the specific needs and goals of the organization, making it both flexible and adaptable.

The data analysis layer encompasses various key activities, which serve as examples but are not exhaustive of all possible analytical operations. Some of the traditional structured data activities include Database Stored Procedures and Data Marts. As big data evolves, the range of analytical activities expands beyond traditional structured databases to include NoSQL Analysis, Predictive Analysis and Machine Learning and Study.

Flexibility and Dynamism

One of the key strengths of the data analysis layer is its flexibility. The methodologies and tools used for analysis can be dynamically adjusted to meet changing business needs. For instance, as new data sources become available or as business objectives evolve, the analysis techniques and tools can be updated to incorporate these changes. This ensures that the analytical processes remain relevant and effective.

Creating a Feedback Loop

A critical aspect of the data analysis layer is the feedback loop it creates within the data ecosystem. The results of the analyses are not merely endpoints; they are fed back into the storage layer. This feedback loop serves several purposes: 1. Continuously improving data models and methods. 2. Adding context and value to the existing data through enriched results.

4.5 Data Utilization and Presentation Layer

Finally, the results from the analysis layer are utilized in various analysis products and presented to the audience. This layer plays a pivotal role in transforming data insights into actionable business strategies and decisions. It ensures that data-driven insights are effectively communicated and leveraged for organizational growth and improvement.

By following this structured approach, non-tech organizations can implement a flexible and reusable data ecosystem incrementally. This method not only minimizes upfront costs and disruptions but also allows for continuous adaptation to changing business needs and technological advancements.

5. Conclusion

This study aims to propose a new framework. The framework outlines the journey of data from acquisition to analysis within organizations, emphasizing flexibility and reusability. The components of the conceptual model include model planning, building a domain conceptual model, and identifying key elements such as data entities, infrastructure layers, and data flows.

The data analysis layer, a critical component of the data ecosystem, transforms raw data into actionable insights through various activities like database stored procedures, data marts, NoSQL analysis, predictive analysis, and machine learning. The adaptability of this layer ensures that it can meet changing business requirements, creating a feedback loop that continuously improves data models and enriches datasets.

This method minimizes upfront costs and disruptions while allowing for continuous adaptation to technological advancements and evolving business needs. By adopting such a data ecosystem, non-tech organizations can enhance their decision-making processes, optimize resource management, and unlock new avenues for growth and innovation in today's dynamic business landscape.

References

[1] Mayhew H, Saleh T, Williams S (2016) Making data analytics work for you—instead of the other way around. McKinsey Q Oct 4, pp 29–41

[2] Loebbecke C, Picot A (2015) Reflections on societal and business model transformation arising from digitization and big data analytics: a research agenda. J Strateg Inf Syst 24(3):149–157

[3] Riasanow, 2021: Core, intertwined, and ecosystem-specific clusters in platform ecosystems: analyzing similarities in the digital transformation of the automotive, blockchain, financial, insurance and IIoT industry, Electronic Markets, 31, 1, 89-104

[4] McAfee, Andrew 2012: Big Data: The Management Revolution

[5] Tobias Riasanow, 2017, DIGITAL TRANSFORMATION IN THE AUTOMOTIVE INDUSTRY: TOWARDS A GENERIC VALUE NETWORK

[6] Andreas Hein, 2020, Digital platform ecosystems, Electron Markets, 30, 1, 87, 98

[7] 2015 Service innovation: A servicedominant logic perspective. MIS Quarterly, 39(1), 155–175.

[8] Tsujimoto et al. 2017, Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies

[9] Kogalovsky, 2009: Conceptual and ontological modeling in information systems, Programming and Computer Software, 35, 5, 241-256

[10] Robinson, 2010: Conceptual modelling: Who needs it. SCS M&S Magazine, 2, 1–7.

[11] Yiming Lin, Hongzhi Wang, Jianzhong Li, Hong Gao 2019: Data source selection for information integration in big data era, Information Sciences, 479, 197-213

Author Profile: Han Sun(1989.3.31-), Female, Han, Shandong, Education: Master's degree, Intermediate title, Major: Advanced Information Systems, Research direction: Data Analysis.