Cuban Journal of Public and Business Administration
ISSN 2664-0856 RNPS 2458 / Vol. 6 No. 3 / September-December (2022) / e249
Available in: https://apye.esceg.cu/index.php/apye/article/view/249

Original article

DATA ECOSYSTEMS REFERENCE ARCHITECTURE BASED ON DATA MESH & DATA FABRIC

ARQUITECTURA DE REFERENCIA DE ECOSISTEMAS DE DATOS BASADA EN DATA MESH & DATA FABRIC

Tatiana Delgado Fernández * https://orcid.org/0000-0002-4323-9674

Havana Technological University "José Antonio Echeverría", Havana, Cuba

* Corresponding author: tatiana.delgado@uic.cu

Classification JEL: O32, O33, O39

DOI: https://doi.org/10.5281/zenodo.7338569

Received: 18/08/2022
Accepted: 25/10/2022

Abstract

Digital transformation requires rapid and profound changes to take advantage of technologies and data in order to make decision-making more effective with agility and self-sustainability. The complexity of data in the modern era and the silos that are generated at big scale drive the emergence of new data management models and architectures that focus on the intrinsic characteristics of digital ecosystems, characterized by the strong interrelationships of various actors through along the value chain, the platforms as the basis for interoperating with each other and the co-evolution of data products that emanate from increasingly heterogeneous sources. This article proposes the design of a reference architecture for data ecosystems based on the data architectures that are best supporting data management in this complex scenario: Data Mesh and Data Fabric, and with the use of knowledge graphs for the integration. As a method, an analysis of the most recent literature on data management and architectures is used to extract the principles and architectural components that are used in the design of such a reference architecture. An abstract representation of the reference architecture of data ecosystems is obtained, whose operational model is theoretically verified. It is the starting point for future research that will be directed towards its implementation in real use cases and organizational modeling related to the roles of the actors involved in the ecosystem reflected in the architecture itself.

Keywords: data architecture, ecosystem, Data Mesh, Data Fabric, knowledge graphs

Resumen

La transformación digital exige cambios acelerados y profundos para aprovechar las tecnologías y los datos en función de hacer más eficaz la toma de decisiones con agilidad y autosostenibilidad. La complejidad de los datos en la era moderna y los silos que se generan a gran escala impulsan la emergencia de nuevos modelos y arquitecturas de gestión de datos que se enfocan a las características intrínsecas de los ecosistemas digitales, caracterizados por las fuertes interrelaciones de diversos actores a lo largo de la cadena de valor, las plataformas como base para interoperar entre ellos y la coevolución de los productos de datos que emanan de fuentes cada vez más heterogéneas. Este artículo propone el diseño de una arquitectura de referencia de ecosistemas de datos basada en las arquitecturas de datos que mejor están soportando la gestión de datos en este complejo escenario: *Data Mesh* y *Data Fabric*, y con el empleo de grafos de conocimiento para la integración. Como método se emplea un análisis de la literatura más reciente sobre gestión y arquitecturas de datos para extraer los principios y componentes arquitectónicos que se emplean en el diseño de tal arquitectura de referencia. Se obtiene una representación abstracta de arquitectura de referencia de ecosistemas de datos, cuyo modelo operacional se verifica teóricamente. La misma es el punto de partida de futuras investigaciones que se encaminarán hacia su implementación en casos de uso reales y el modelado organizacional relativo a los roles de los actores que se involucran en el ecosistema reflejado en la propia arquitectura.

Palabras clave: arquitectura de datos, ecosistema, Data Mesh, Data Fabric, grafos de conocimientos

Introduction

Data has an increasingly important role and value in facilitating decision-making. The volume, variety, speed, veracity, as a quality requirement, and the value of data, are usually used to define the Big data concept, a term that, more than large or massive data, characterizes its complexity and the paradigmatic change that has been occurring in the architectures that manage them.

So-called analytics data is becoming an increasingly critical component of the technology landscape. They are the basis for visualizations and reports that provide information about a business or organization. Additionally, they are used to manage machine learning models that augment the business with data-driven intelligence, which is a key factor for organizations to move from intuition and instinct-driven decision-making to acting based on observations and data-supported predictions. It enables a technological shift from rule-based algorithms designed by humans to machine learning models.[1]

In this new scenario, the challenge of "data silos" becomes more palpable due to their increasingly heterogeneous nature. A data silo means that data is not as accessible as it should be, or perhaps not to teams other than the ones generating it. If a significant amount of time is required to decode the data so that it is translatable to another team, there are likely one or more data silos in the organization. Data silos arise from structural (many layers of separation between groups), cultural (i.e. keeping data separate, rather than working together), and technological (applications may not be designed to integrate) issues.[2]

The models and architectures that support data management are changing to meet these challenges. One of the most mentioned architectures in advanced communities in this context is Data Mesh or data fabric, considered a decentralized socio-technical approach to share, access, and manage analytical data in complex and large-scale environments, within or between organizations. It is based on four fundamental principles: domain ownership, data as a product, data self-service platform, and federated computing governance.[1]

Another emerging architecture making its way onto the data management scene is the Data Fabric (DF), which can broadly be defined as a set of data management principles, guiding practices, communities, and standards that can "...optimize access to an organization's distributed data and intelligently curate and organize it for self-service delivery."[3] It is a system that provides a unified architecture for managing and serving data. They are generally realized as service-oriented distributed systems where the sets of services provide consistent interfaces and mechanisms for accessing data and storage capabilities.[4]

Considering the diverse and multiple interrelationships that arise between data sets from different domains and the different actors that manage them, these scenarios are frequently described as digital ecosystems,[5] within which data ecosystems emerge with a strong force, given the crucial importance of integrated data management to make decisions more effectively and based on context.

A data ecosystem can be defined as a set of networks composed of autonomous actors that directly or indirectly consume, produce, or provide data and other related resources (for example, software, services, and infrastructure). Each actor performs one or more roles and is connected to other actors through their relations in such a way that the collaboration and competition of the actors promote the self-regulation of the data ecosystem.[6]

The objective of this article is to propose a data ecosystem reference architecture based on the fusion of components of the emerging Data Mesh and Data Fabric architectures, to offer an abstract model from which different data ecosystem architectures can be instantiated, from various domains, companies, and even, to eliminate inter-agency data silos, and achieve a central level of government data integration, as part of the implementation of digital transformation policies.

Materials and Methods

To define the reference architecture of data ecosystems, a hybrid research methodology is used where literature analysis methods are integrated to scientifically support the proposal, together with other methods of modeling modern data architectures, which respond to a management effective management of complex data generated in the era of digital transformation. **Figure 1** shows the scheme of the phases that methodologically guided the development of the research.

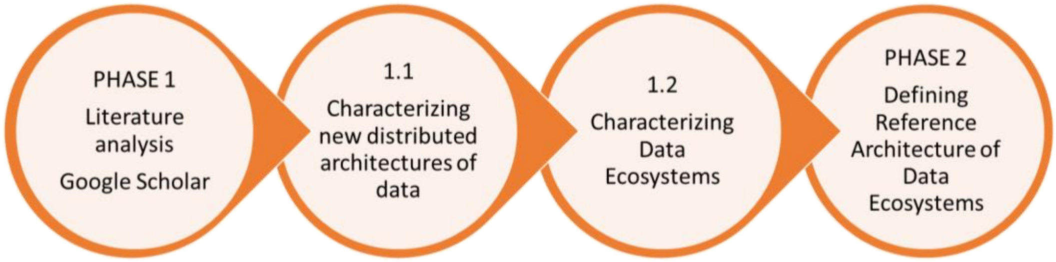

Figure 1. Methodology to define the reference architecture of data ecosystems
Source: self-made

The first phase corresponds to the literature analysis, for which Google Scholar was used, due to its versatility, covering a wide variety of publications, such as articles, books, conference proceedings, theses, and other materials. Also included are some sources of the so-called gray literature, in this case, electronic sources from global leaders in the field of data management, which are setting standards in data architectures, particularly about emerging Data architectures. Mesh and Data Fabric. The objective of applying this method is to reveal the principles and distinctive characteristics of these architectures that, because they are so disruptive, there is little scientific evidence of their implementation. However, they are identified as trends in reports from global consultancies, such as Gartner, and vendors, such as IBM and Microsoft. Once the architectures are analyzed, whose components will be evaluated for reuse in the reference architecture, the conceptual framework of data ecosystems is established, which occupies another body of knowledge, although highly interrelated with the first one.

The second phase is the design of the data ecosystem reference architecture. With the architectural components of Data Mesh and Data Fabric, which resulted from the study of the selected architectures in the first phase, the new reference architecture is designed, paying careful attention to the principles inherited from its predecessors and the data ecosystems themselves.

Results

Data mesh

The four principles underlying the logical architecture and operating model of a data fabric are (1) domain-oriented decentralized data ownership, (2) data as a product, (3) self-service data platform, and (4) federated computational governance.[1] These principles are described below:

1. Domain ownership principle. It decentralizes ownership of the analytics data to the business domains closest to the data, be it the source of the data or its primary consumers. It decomposes (analytics) data logically, based on the business domain it represents, and manages the domain-

oriented data lifecycle independently. Architecturally and organizationally aligns business, technological and analytical data. There are three archetypes of domain-oriented data: source-aligned domain data, aggregated domain data, and analytics data.

2. Principle of data as a product. With this principle in place, domain-oriented data is shared as a product directly with data users: data analysts, data scientists, etc. Each data product is self-contained; its lifecycle and model are managed independently of the others. Data as a Product introduces a new logical architecture unit called the data quantum, which controls and encapsulates all the structural components needed to share data as a product (data, metadata, code, policy, and declaration of infrastructure dependencies) in an autonomous way. Get more value from data by sharing and using data across organizational boundaries.
3. Principle of the data self-service platform. This principle leads to a new generation of self-service data platform services that enable cross-functional teams across domains to share data. The platform's services are focused on removing friction from the end-to-end journey of data exchange, from source to consumer. Platform services manage the complete life cycle of individual data products. They handle a trusted fabric of interconnected data products. They provide fabric-level experiences, such as showing the emergent knowledge graph and lineage across the mesh. The platform streamlines the experience for data users to discover, access, and use data products. It also streamlines the experience of the data providers to create, deploy, and maintain data products.
4. Principle of federated computational governance. This principle creates a data governance operating model based on a federated decision-making and accountability structure, with a team comprised of domain representatives, data platform, and subject matter experts: legal, compliance, security, etc. The operating model creates an incentive and accountability structure that balances domain autonomy and agility with global fabric interoperability. The governance execution model relies heavily on codifying and automating policies at a granular level, for each data product, through platform services.

The data mesh provides the flexible and resilient integration of data sources across platforms and business users, making them available wherever they are needed and depending on where they are hosted.[7]

The multi-plane platform[1] of the data fabric allows one to distinguish between different platform services based on their scope of operation without imposing strict stratification. The three deck planes include the following:

- Data infrastructure plane. Atomic services to provision and manage physical resources such as storage, pipeline orchestration, computing, etc.
- Plane of experience of the data product. Higher-level abstraction services operate directly with a data product and enable data product producers and consumers to create, access, and protect a data product, among other operations that run on a data product.

- Mesh experience plane. Services that operate on a fabric of interconnected data products, such as searching for data products and observing the lineage of data between them.

These blueprints can be directly accessed by platform consumers (data product developers, consumers, owners, government function, etc.).

Figure 2 offers a view of data mesh/data fabric as an integrated layer of connected data.

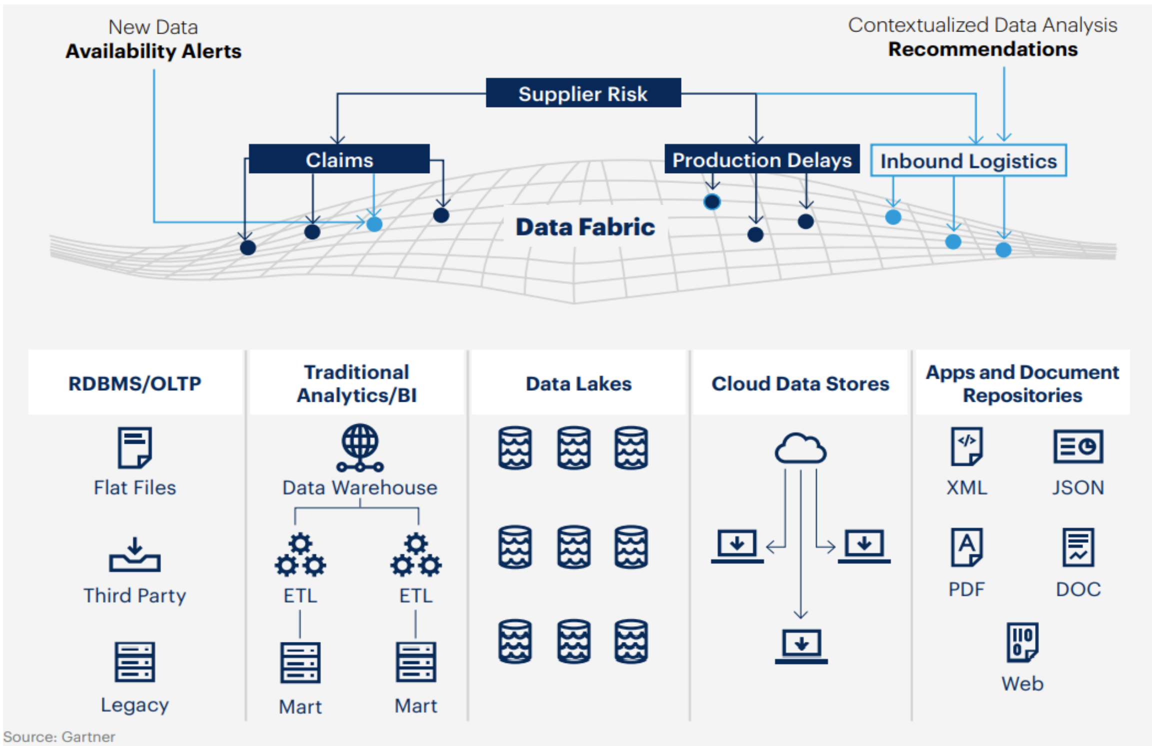

Figure 2. Data fabric as an integrated layer of connected data
Source: Gartner[7]

This Gartner approach to the Data Mesh is closer to that of the Data Fabric architecture, because it more consciously identifies this integrated technology layer, which is generally resolved in the form of knowledge graphs.

Data Fabric

Data Fabric, which in Spanish can also be translated as Data Mesh and, in order not to be confused with the previous approach, will be called by its term or acronym in English (DF), constitutes an information architecture and a platform for data management and integration and provides interfaces, APIs, and services for the systems integration and communication involved. From a systems-level perspective, DF can be viewed as a communication substrate that provides a unified mechanism for data access and manipulation for all project tools.[4]. This data architecture is more of a framework that enables automatic and intelligent end-to-end deployment of multiple data channels as well as cloud environments.[3]

The distributed nature of DFs allows for scalability, flexible deployment, and system adaptation. It is often leveraged to, for example, facilitate system integration across organizational boundaries or combine the use of local and cloud-based resources. While primarily designed as substrates for data management and system-to-system communication, DFs can also expose interfaces and tools to end users to facilitate the development of mechanisms to manage, search, and analyze data.[4]

Since an organization's distributed data will evolve, both in content and in scale and format, it is essential to have a flexible and scalable approach. These concepts of evolution and scalability are fundamental in Data Fabric architectures. A DF model should always strive to cooperate with communities and workgroups due to the inherent value of the connected data and services. Some of that cooperation can result in formally published standards and approaches. There are also design principles to consider, such as:

- Availability of data refers to the fact that the data must be intuitive.
- Value of the data, that is, it has an intrinsic value.
- Interconnected data, which means data is inherently more valuable when connected.
- The data must be FAIR, known by its acronym in English means findable, accessible, interoperable, and reusable.
- Harmony between data, services, and software. It is advisable to use ontologies and other semantically expressive approaches, both technically and conceptually.
- Learning from your data, meaning if captured and connected correctly, you can learn a lot about data and provide value to the overall effort and mission. It is recommended to use artificial intelligence to learn patterns that could be useful at a broader level, statistical analysis to determine the most efficient ways to solve problems, and predictive workflows based on current knowledge, among other techniques.

To increase data health, the Data Fabric offers integrated data quality, data pre-processing, and information governance capabilities enabled by machine learning and enhanced automation.[8]

DF is summarized as a data management design that enables the integration and exchange of data between heterogeneous sources, to achieve flexible, reusable, and augmented data integration, using knowledge graphs, semantics, and machine learning/artificial intelligence, in active metadata, to support faster, and, in some cases, automated data access and sharing, regardless of deployment options, use cases (operational or analytics), or architectural approaches.[7]

Data-level integration in the Data Fabric frequently occurs through knowledge graphs (KGs). A KG is a conceptual model of a knowledge domain, in this case, your product design and creation process. Domain experts use such a KG to describe and solve domain-related problems, using their real-world concepts, vocabulary, and relationships between these concepts. The KG does not necessarily have to contain all the data available in an organization or the domain it represents. This would be undesirable and usually even unrealistic. Instead, there are several options for relating KG and data, and a KG typically brings these aspects together, balancing flexibility, cost, and performance.[4]

Data Fabric and Data Mesh provide architectures for accessing data across multiple technologies and platforms. They differ in that the former is focused on technology, while the data fabric focuses more on organizational change. They can be mixed to take advantage of both because of the strong compatibility between both of them. Data Fabric makes more sense of how data integrates harmoniously by explicitly providing the knowledge graph variant. This will be exploited in the reference architecture designed in this research.

Data Ecosystems

The metaphor of ecosystems has been used to describe multiple and varied interrelationships between many actors and infrastructures that contribute to the creation of a resource, for example, a business, service, or software.[9] In this sense, the ecosystems presented go beyond the traditional value chains and industrial structure by having three main characteristics: network, platform, and co-evolution. The first characteristic establishes the existence of a loose network of players, including developers, vendors, resellers, and technology and infrastructure providers. All actors are committed to producing value or extracting value from the ecosystem. The second feature is a "platform" (for example, services, tools, or technologies) that ecosystem actors can use to generate benefits. The platform allows different actors to contribute to the ecosystem and results in a set of products or services. Finally, the ecosystem allows actors and data products to evolve together, that is, to be part of an ecosystem that requires collaboration and connection between different actors in different fields of specialization and knowledge, and between the artifacts, they generate. At the same time, being part of the ecosystem allows actors to have access to each other, as providers, innovators, or problem solvers, whether they work independently or within research organizations, private or public organizations.[6]

Thus, a data ecosystem can be viewed as another instance of a digital ecosystem.[5] Furthermore, a data ecosystem can be thought of as part of multiple types of ecosystems organized around companies, resources, and products provided by different actors. The broader objectives of innovation and value creation are translated into more specific terms related to each specific ecosystem context. In particular, the data can be used to support business, drive innovation, promote transparency for governments, validate research, and many other purposes. In addition to being interconnected, the boundaries between a data ecosystem and other ecosystems are difficult to define. For example, a data ecosystem may imply software ecosystems over the network of actors involved in the development and supply of data-related software.[9] In public administration, examples of data ecosystems for government arise as well.[10]

The frameworks and architectures of data ecosystems are still in their infancy and although there are some interesting proposals,[11-14] there is much room for research in search of an optimized management model for complex data, in distributed environments, such as data systems. socio-technical, and in an interoperable and integrated manner.

Data management mechanisms and architectures are taxonomic components of digital transformation,[15] facilitating its adoption. In this article, data ecosystems are placed at the center to propose an architecture that is capable of supporting the technological and organizational implementations of this paradigm, whose very essence is based on digital ecosystems.

Data ecosystem architecture based on Data Mesh and Data Fabric

To design the data ecosystem reference architecture, the four principles of Data Mesh1 are reused: domain ownership, data as a product, data self-service platform, and federated computational governance. **Figure 3** presents the proposed Data Ecosystem Reference Architecture.

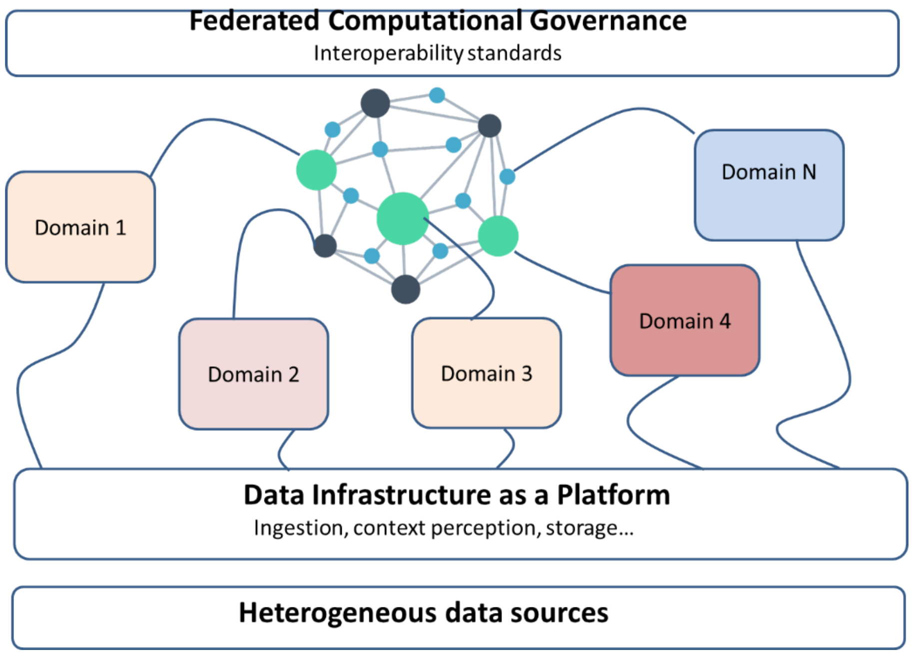

Figure 3. Reference Architecture of Data Ecosystem
Source: self-made

Knowledge graphs (KG) are assumed as the ideal architectural component to achieve integration between the different data products served in the domains.

In general, the instances of the concepts of a graph are considered the "data" of the KG. However, that does not mean that the KG should contain all the data available in an organization. Instead, there are several options for relating KG and data, balancing flexibility, cost, and performance.4 These options include:

- KG materialized with direct individuals. A KG can contain all of its individuals within its infrastructure. This is the traditional way to build a KG.16-19
- Virtual KG: While a KG may contain individuals by concept, it is not required to hold them physically in all cases. A virtual KG dynamically obtains some of its individuals from other storage types and provides them to the user who requests them. This makes them scale better when the number of individuals increases.20-21
- Reference KG: Sometimes, storing individuals in the KG is not necessary. Instead, their access through the KG is enough. The sheer size of the related data may be prohibitive, or a KG may not be adequate to represent their individuals. This is, for example, applicable to time series data, which is very common in project production scenarios. However, a KG does not have the proper structure to efficiently handle massive amounts of time-based measurements for many operational physical properties. Instead, the KG should point to the appropriate data access entry in the Data Fabric.

The construction of the graph is associated with some of these variants, but also with the nature of the data product of each domain. In the literature, there are detailed methodologies for knowledge graph construction.[17] Likewise, implementations in specific domains are presented, such as, for example, a study carried out for the graphs´ design in a scenario related to epidemiological analyzes of COVID 19.[16] Other examples, in the case of business knowledge graphs, are being addressed more frequently in the direct mode[18-19] and as virtual KGs.[20-21] However, in none of these cases are the infrastructure and domain layers included, with the delivery of data products by domain, as shown in the proposed architecture.

Discussion

The data ecosystem reference architecture is verified by verifying that it integrates the three basic characteristics of the ecosystems, described above: network, platform and coevolution, and also by analyzing how it reflects the architectural components of the architectures. of data that give rise to it: Data Mesh and Data Fabric.

The characteristics present in an ecosystem are explicitly conceived, first, in its operation as a network, with the KG architectural component provided by the Data Fabric architecture model, which allows linking the data and its properties without considering the applications of each domain, but rather using the data as a product; while its platform nature is reflected in the incorporation of the infrastructure layer, which follows the self-serving platform philosophy of Data Mesh.

The data infrastructure layer as a platform is responsible for ingesting data from heterogeneous sources, using ETL-extract, transform and load (for transactional data that will be managed in data warehouse domains), or ELT -extract, load and transform (in the case of complex data to which management mechanisms based on Big Data are applied) or any other mechanism aligned to the sources.

The domains are responsible for generating the data product, as defined in the Data Mesh blueprint architecture.[1] The fabric experience plane optimizes the experience of the people who need to operate, govern, and query the fabric as a whole. In the case of the data ecosystem reference architecture presented, the specific capacity of the data services offered by each domain is used to serve the graphs that link the data at an integrated level, under the Data Fabric philosophy. At the same time, the policies and standards recommended in the federated computational governance layer are considered. For example, members of the governance team and data product owners, working within domains, interact with services in this plane to assess the current state of policies, monitor the overall operational health of the fabric, and search for existing data products.1 It is also used by consumers and providers of data products in scenarios where they need to work with a collection of data products, such as data search and retrieval. For the proposed reference architecture, the integrated data is consulted through the tools associated with knowledge graphs, using SPARQL or other applications that embed their functionality.

The data product experience plane is optimized for the delivery and consumption of data products via APIs and through knowledge graphs when integrating multiple data from different domains.

Theoretical implications of the research

Data ecosystem architecture, like Data Mesh, calls for a fundamental change in the way analytics data is managed, used, and consumed, as described:

- The decentralized data ownership model pushes data ownership and responsibility to the business domains from where the data is produced or used, prioritizing a federated model with computational policies embedded in the nodes of the fabric.
- Data is served as products, which take better advantage of the intrinsic characteristics of each data source, while being served in the way that best satisfies the consumer experience.
- Architecturally, it moves from collecting data in monolithic warehouses and lakes to connecting data through a distributed fabric of data products accessed through standardized protocols, while technologically, solutions treat the data and the code that supports it. maintained as an active autonomous unit.

This article, whose main contribution is the very design of a data ecosystem reference architecture that is verified with respect to the principles of those architectures that give rise to it, also has limitations. The most important of these is that it is based on a theoretical approach, so future research should address empirical and case-based methods that implement the reference architecture for data ecosystems in real domains. On the other hand, it is necessary to continue investigating the roles of the actors that involve a data ecosystem and reflect it in the architecture so that it can be implemented in a governable way**.**

Conclusion

Data ecosystems are being ratified as the best model to represent the multiple interconnections between different actors that result in a set of products or services that are generated from the also interconnected data. Opposed to data silos, data ecosystems ensure interoperability and integration at the data level.

They must be built on flexible and scalable architectures that facilitate automatic and intelligent end-to-end deployment of multiple data channels.

Trends around frameworks for building such architectures point to Data Mesh/Data Fabric. The data ecosystem reference architecture presented at an abstract level in this article inherits the principles and characteristics of these frameworks to contemplate a decentralized data ownership model, where business domains are concerned with providing data products to the ecosystem, for the consumption of any other actor/domain or to be integrated into the fabric (graph) from where they are consumed holistically contextualized.

Future work in this direction will be aimed at carrying out proofs of concept that validate the reference architecture and incorporating organizational analysis and the practical implications of its adoption.

Bibliographic references

1. Dehghani Z. Data Mesh: Delivering Data-Driven Value at Scale (1.ed - preview version), O'Reilly Media, Inc. 2022. [Consulted 5 september 2022]. Available in: https://www.oreilly.com/library/view/data-mesh/9781492092384/.
2. Fortney J, McDonnell M, Johnson D, Chalk S. Data Fabric and Data as a" First Class Citizen"; 2022. [Consulted 1 september 2022]. Available in: http://dx.doi.org/10.13140/RG.2.2.14510.18240
3. IBM, "Data fabric," 2021. [Online]. [Consulted 1 september 2022] Available: https://www.ibm.com/analytics/data-fabric
4. Östberg PO, Vyhmeister E, Castañé GG, Meyers B, Van Noten J. Domain Models and Data Modeling as Drivers for Data Management: The ASSISTANT Data Fabric Approach. IFAC-PapersOnLine. 2022 Jan 1;55(10):19-24. [Consulted 4 september 2022]. Available in: https://doi.org/10.1016/j.ifacol.2022.09.362
5. Delgado T. Una arquitectura de Ecosistemas de Datos Espaciales. XVI Convención y Feria INFORMATICA 2016: Conectando sociedades 2016; 1-6. ISBN 978-959-289-122-7.
6. de Oliveira EF, Silveira MS. Open government data in Brazil a systematic review of its uses and issues. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age 2018 May 30:1-9. [Consulted 21 august 2022]. Available in: https://doi.org/10.1145/3209281.3209335).
7. Gartner. Understand the role of Data Fabric. Guides for Effective Business Decision Making; 2022. [Consulted 21 august 2022]. Available in: https://www.gartner.com/en/publications/essential-guide-to-data-fabric.
8. Liu CM, Badigineni M, Lu SW. Adaptive Blocksize for IoT Payload Data on Fabric Blockchain. In2021 30th Wireless and Optical Communications Conference (WOCC) IEEE. 2021 Oct; 7: 92-96). [Consulted 2 august 2022]. Available in: http://doi.org/10.1109/WOCC53213.2021.9602935.
9. Farias VG, Santos R, Wiese I, Serebrenik A, Constantinou E. Investigating Power Relations in Open Source Software Ecosystems. InAnais Estendidos do XII Congresso Brasileiro de Software: Teoria e Prática 2021 Sep 27 (pp. 53-59). SBC. [Consulted 23 july 2022]. Available in: https://doi.org/10.5753/cbsoft_estendido.2021.17282
10. Shah SI, Peristeras V, Magnisalis I. Government big data ecosystem: definitions, types of data, actors, and roles and the impact in public administrations. ACM Journal of Data and Information

Quality. 2021 May 6;13(2):1-25. [Consulted 13 august 2022]. Available in: https://doi.org/10.1145/3425709

11. Hernandez-Almazan JA, Chalmeta R, Roque-Hernández RV, Machucho-Cadena R. A Framework to Build a Big Data Ecosystem Oriented to the Collaborative Networked Organization. Applied Sciences. 2022 12;12(22):11494. [Consulted 5 november 2022]. Available in: https://doi.org/10.3390/ app122211494.
12. Herrera F, Sosa R, Delgado T. GeoBI and big VGI for crime analysis and report. In2015 3rd International Conference on Future Internet of Things and Cloud 2015 Aug 24 (pp. 481-488). IEEE. [Consulted 12 july 2022]. Available in: https://doi.org/10.1109/FiCloud.2015.112
13. Orenga-Roglá S, Chalmeta R. Framework for implementing a big data ecosystem in organizations. Communications of the ACM. 2018 Dec 19;62(1):58-65. [Consulted 21 july 2022]. Available in: https://doi.org/10.1145/3210752
14. Singh KN, Behera RK, Mantri JK. Big data ecosystem: review on architectural evolution. Emerging Technologies in Data Mining and Information Security. 2019:335-45. [Consulted 1 august 2022]. Available in: https://doi.org/10.1007/978-981-13-1498-8_30
15. Fernández TD. Taxonomía de transformación digital. Revista Cubana de transformación digital. 2020;1(1):4-23. [Consulted 15 august 2022]. Available in: https://rctd.uic.cu/rctd/article/view/62.
16. Delgado T, Stuart ML, Delgado M. Grafos de conocimiento para gestionar información epidemiológica sobre COVID-19. Revista Cubana de Información en Ciencias de la Salud. 2021 Dec;32(4). [Consulted 12 august 2022]. Available in: http://rcics.sld.cu/index.php/acimed/article/view/1686.
17. Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutierrez C, Gayo JE, Kirrane S, Neumaier S, Polleres A, Navigli R. Knowledge graphs. arXiv preprint arXiv:2003.02320. 2020; Mar 4. [Consulted 2 august 2022]. Available in: https://arxiv.org/abs/2003.02320.
18. Gomez-Perez JM, Pan JZ, Vetere G, Wu H. Enterprise knowledge graph: An introduction. InExploiting linked data and knowledge graphs in large organisations 2017 (pp. 1-14). Springer, Cham. [Consulted 20 july 2022]. Available in: https:/doi.org/10.1007/978-3-319-45654-6_1.
19. Sequeda J, Lassila O. Designing and building enterprise knowledge graphs. Synthesis Lectures on Data, Semantics, and Knowledge. 2021 Aug 3;11(1):1-65. [Consulted 3 august 2022]. Available in: https://doi.org/10.2200/S01105ED1V01Y202105DSK020.
20. Cárdenas ML, Fernández TD, Fernández MD, de la Iglesia Campos M. Grafos virtuales de conocimiento para la integración de datos empresariales en una empresa cubana. Revista Cubana de Administración Pública y Empresarial. 2022 Apr 20;6(1):e211. [Consulted 14 july 2022]. Available in: https://doi.org/10.5281/zenodo.6472957.
21. Xiao G, Ding L, Cogrel B, Calvanese D. Virtual knowledge graphs: An overview of systems and use cases. Data Intelligence. 2019 Jun 1;1(3):201-23. [Consulted 2 august 2022]. Available in: https://doi.org/10.1162/dint_a_00011

Conflict of interests

The author declares that she has no conflicts of interest