Educational Administration: Theory and Practice
2024, 30(6), 1882-1886
ISSN: 2148-2403
https://kuey.net/
Research Article

Data Architecture Development Through Data Metrics Using Investment Banking Sector Insights

Wiktoria Gromowa-Cieslik*

*Faculty of Management, Wroclaw University of Science and Technology. victory.gromova@gmail.com

Citation: Wiktoria Gromowa-Cieslik, et al, (2024) Data Architecture Development Through Data Metrics Using Investment Banking Sector Insights, *Educational Administration: Theory And Practice*, *30*(6), 1882-1886,
Doi: 10.53555/kuey.v30i6.5610

ARTICLE INFO **ABSTRACT**

The topic of data architecture, which is inherently linked to the broad concept of data, remains relevant and requires further analysis and development despite the evolving methods and models used among practitioners. This is most easily observed in the example of a specific area, namely the investment banking sector, which has been chosen for analysis. The analysis confirms, in addition to affirming the above statement, that creating an effective data architecture should somehow consider not only the business objectives of the chosen area or department, but also the strategy for managing the organizational structure. Thus, to combine business needs with other available information about these needs, the creation of data metrics is proposed. This model is based on principles of matching current data to the created structure in space. In the considered example, the structure assumes a cubic form and is modeled using matrix transformation. Matching data to this structure is an optimization task aimed to be performed with computer software. The proposed model has practical application in the selected department, and further case studies are being conducted to assess the scalability of the proposed approach.

Index Terms—Data architecture, data metrics, investment banking

I. INTRODUCTION

As an introduction to the topics that will be covered in the article (data architecture and data metrics used to its development), it is worth starting with defining the specifics of how investment banks operate. Investment banks are often large organizations, employing tens of thousands of people in various countries around the world. Depending on the bank's strategy, which defines the development of specific business lines, or cooperation with clients according to established goals, the organizational structure changes. For example, being a leader in American consumer and social banking, JP Morgan Chase & Co. has over 50% of its employees in America. However, we often see different situations as well. According to the latest data from the end of 2023, private banking, dedicated to wealthy clients, is concentrated in North America by 32%. And even if Bank of America is the leader in this region, it is not the leader in every area of private banking, which means that the competencies needed to serve it are not concentrated in one organization and in one region. This creates healthy competition in the local job market, but also imposes certain limitations. Therefore, banks decide to make the most of the opportunity to create a global organization [1], which, on the one hand, increases their revenues, but on the other hand, enables revenue growth due to global diversification. In this process, digitization plays a major role, although it is not always an effective solution for all cases. There are still regulations that are specific to a group of countries or even to individual countries, which necessitates the separation of certain processes to manage them in accordance with the requirements of the financial regulator. Hence, it can be concluded that not only the strategy of banks influences the shaping of the organizational structure, but also aspects such as access to talents fitting for business and technical areas, economic potential of the market, access to technological resources, and the geographical specifics of the location (for example, in case of data center creation).

The aforementioned aspect of digitization, despite certain limitations, is one of the key topics in the financial sector [2]. After each major crisis in this sector, new regulations have emerged to prevent their recurrence (e.g., basic Basel principles or principles concerning the financial market infrastructure (Principles for financial market infrastructures – PFMI). The main goals of these principles remain the following: quick access to financial information, transparency in resource management, and adequate risk management. Therefore, digitization, in one of its aspects, helps in accessing data by ensuring the ability to view information in real-time requires the use of appropriate IT resources. On the other hand, since all activities and business operations, especially those in banks, take place in a digital environment, the need to ensure cybersecurity leads towards cloud solutions [3]. It is estimated that by 2025, servicing major banking systems in the cloud will count for 14% of total IT spending. Already now, reports show that in 2023, spending on cloud solutions in the financial sector alone increased by 17% compared to 2022. The observed trend indicates that the movement towards, for the time being, private cloud is embedded in firms' strategies, and following it is inevitable. However, in the case of large investment banks, it is more complicated because they still rely on systems developed decades ago, on older technologies that, at best, require rewriting applications to new technology, but most often the entire area requires a thorough overhaul. To make this possible, knowledge of the actual state of the systemic architecture is needed, meaning the description of the structure of IT systems and the relationships between their components.

In both scientific and business literature, much is said about system architecture and the role of system architects [5] in organizations. Key topics in this area include: integration of IT systems, selection of resources and tools, setting standards, and organizational development strategies in relation to business requirements. All these aspects influence each other - one cannot focus only on one of them, leaving the rest neglected. If, for example, the goal is to integrate IT systems, which may even arise from business needs, it will certainly not be possible without developing appropriate methodologies for achieving this goal, but also to ensure that the system can operate effectively after implementation. Additionally, a well-designed system architecture contributes to effective management of IT networks, ensures access to information under security conditions, and ensures compliance with security principles [4]. Since banking relies on system architecture that was often established decades ago or simply on systems that connect without a properly thought-out and described schema, the issue of proper selection both IT and non-IT resources is one of the most important, especially when creating a long-term strategy. The rapid pace of technological development intensifies the pressure to create a thoughtful IT foundation (for example, when transitioning to cloud solutions), which is flexible enough to adapt to the necessity of modifying established standards, protocols, but also ensures continuity of business processes during the transition period.

II. BASIC MODELS OF DATA ARCHITECTURE

In addition to the focus on implementing modern technological solutions for improved management of processes and resources, it is still only a tool for an important aspect of organizational functioning - achieving financial goals, which is possible, among other things, by using information. In reality, conducting any business activity (even production or construction) is based on having data, which allows assessment of the competitive market, customer needs, or industry development direction. Hence, it is clear that data is an integral part of the information system, which consists of both the aforementioned system architecture and data architecture, defined as a set of models, policies, rules, and standards used in the process of managing the way data is collected, stored, integrated, and used in the enterprise. In the field of data architecture, the literature often discusses several models - Zachman's Enterprise Framework and TOGAF (The Open Group Architecture Framework). Zachman's model, basically, focuses on organizing architecture artifacts (data, connections, people, time, and goals), taking into account individual levels - corporate roles of people involved in the process. TOGAF, on the other hand, primarily considers the strategic goals and tasks of the organization, business processes in an integrated coherent organizational architecture [2]. In practice, attempts have been made to apply the BIAN (Banking Industry Architecture Network) model or the COBIT standard (Control Objectives for Information and related Technology) or Data Mesh (a concept where business or technical units are responsible for data from their area). There are also initiatives within Data Governance - a concept where key areas are: responsibility for completeness, correctness, timeliness, consistency, and security of data; building awareness of data standards and tools (e.g., Collibra Data Governance or internally developed tools based on agent-based solutions). What connects these concepts and models, in addition to the common goal - arranging business processes, data, and IT systems in an orderly manner, is also the need for top-down (with the support of organizational management) and comprehensive (wide range of changes) approach to the topic, which often becomes the cause of failure. According to 2023 statistics, nearly half of the companies that make changes in Data Governance fail in this area, precisely due to the reasons mentioned above. Despite attempts to organize information flow, companies still incur significant costs (according to research conducted by Gartner in 2022, around $12.9 million) due to poor data quality.

III. DATA METRICS AND MODEL FOR DATA ARCHITECTURE DEVELOPMENT

Bearing in mind the above considerations, the objective of this paragraph is to present an alternative model that focuses on addressing the research problem related to the lack of a defined data management method allowing for effective organization of data access from the bottom up, i.e., by defining and implementing internal goals of selected organizational units, in this case – investment banks. Data metrics [6] [7] will be used for this purpose, allowing for the construction of a mathematically defined and scalable data structure. In literature, the concept of metrics is most commonly used in data quality assessment or as indicators calculated based on adopted parameters [8]. In this work, however, the concept of data metrics is defined as a data or data array referring to a specific business need and including basic information about the data or data array. Let us consider the example of the following business need: "customer service quality control." To effectively address this need, eight areas are taken into consideration: client issues, customer meetings, service reviews, account reviews, customer interactions (other than meetings), log information, customer service team information, and strategic customer service information. Data metrics are created in each of these areas. The business assumption is that each metric from a given area directly results from the preceding one, creating a linear dependency of points in space. Let's consider the example of "Client Issues" (Figure 1).

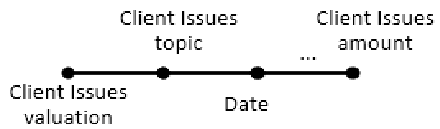

Figure 1: Client Issues linear dependency example

If we designate this area as A, then "Client Issues valuation" will be A1, "Client Issues topic" - A2, "Date" - A3...."Client Issues amount" - An. It is not permissible to create two or more points emerging parallelly (Figure 2).

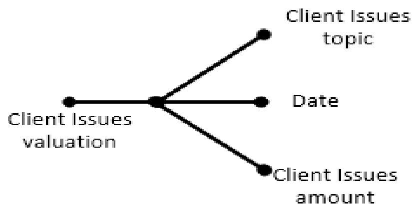

Figure 2: Client Issues parallel dependency example

Each of the subsequent areas will be labeled as B ("customer meetings") and each subsequent metric as B1, B2, ...Bn; C ("service reviews") with metrics C1, C2,...Cn; D ("account reviews") with metrics D1, D2,...Dn; E ("customer interactions (other than meetings)") with metrics E1, E2,...En; F ("log information") with metrics F1, F2,...Fn; G ("customer service team information") with metrics G1, G2,...Gn; H ("strategic customer service information") with metrics H1, H2,...Hn.

The model's assumption is that each layer of data metrics from each area forms a cube (Figure 3). By layer is meant a set of data metrics X1 (A1, B1, C1, D1, E1, F1, G1, H1). Each subsequent layer consists of X2 (A2, B2, C2, D2, E2, F2, G2, H2) ... Xn (An, Bn, Cn, Dn, En, Fn, Gn, Hn). Each cube illustrates the structure of data metrics related to a specific business matter. Structures may overlap to utilize the same data.

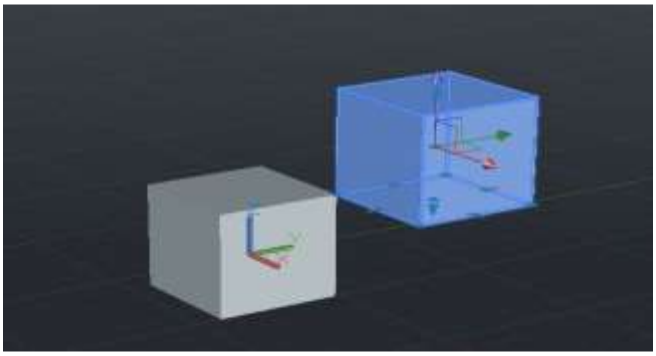

Figure 3: Cubes formed by data metrics from different layers related to specific business matter (AutoCAD)

The description above outlines how data is shaped according to business needs. However, the definition also mentions that basic information about the data or data array should be taken into account. Using the example provided, three groups that will ultimately form a matrix can be listed as follows:

$$S1 = \begin{bmatrix} s11 & s12 & s13 \\ s21 & s22 & s23 \\ s31 & s32 & s33 \end{bmatrix},$$

$$S2 = \begin{bmatrix} s11 & s12 & s13 \\ s21 & s22 & s23 \\ s31 & s32 & s33 \end{bmatrix},...$$

$$Sn = \begin{bmatrix} s11 & s12 & s13 \\ s21 & s22 & s23 \\ s31 & s32 & s33 \end{bmatrix}$$

Basic system information:
s11 - system of the selected department/another department's system (selection 0 or 1)
s12 - visualization product used by the selected department or another department's system (selection 0 or 1)
s13 - whether the data is in the CRM (consolidated database) (no = 0, yes = 1)

Basic data-related information:
s21 - whether the data format is adapted to the visualization product (no = 0, yes = 1)
s22 - whether data classification includes confidential data (no = 0, yes = 1)
s23 - whether the system is compliant with the above data classification (no = 0, yes = 1)

Basic compliance information:
s31 - whether data is registered in the collective database (no = 0, yes = 1)
s32 - whether a change management process is established (no = 0, yes = 1)
s33 - whether changes in cloud/on-prem architecture for data processing/transmission are required or have been made (no = 0, yes = 1)

$$U = U1 \dots U8 = \begin{bmatrix} u \\ v \\ w \end{bmatrix},$$

U – is a 3x1 column matrix describing a vector (point) before transformation, i.e., the vertices of the cube.

The optimization task (using software) is to match data metrics as closely as possible (considering information from matrix S) to the assumed cubic structure:

$$X1 = U * S = \begin{bmatrix} u \\ v \\ w \end{bmatrix} * \begin{bmatrix} s11 & s12 & s13 \\ s21 & s22 & s23 \\ s31 & s32 & s33 \end{bmatrix}$$

If one-to-one matching is not possible, point shifts are permissible. However, each subsequent layer is built from basic structural assumptions, meaning that matrix S is superimposed on the structure of matrix U, which is shifted in space by a fixed unit and forms the vertices of the cube.

Conclusion

In the considered sector of investment banking, the data domain occupies a key position as it provides vital information for making strategic decisions, including financial ones. If the data is not structured at the substantive (managerial) level, it is difficult to expect the effectiveness of maintaining this data at the logical (data architecture) or technical (system architecture) level. These areas are closely connected and require simultaneous consideration. Therefore, in the proposed model, data metrics are created, which take into account the mentioned aspects and allow for the grassroot creation of data structure (architecture), in the example provided, a cubic structure. Each of the vertices of the cube consists of data taken from three informational groups. This model is currently tested on data related to four different business needs from a given department and will be further expanded to include data from additional departments to verify the scalability thesis.

REFERENCES

1. Anjan Roy, “Organization Structure and Risk Taking in Banking,” Risk Management, vol. 10, no. 2, pp. 122-134, April 2008.
2. KPMG International Cooperative, “Managing the data challenge in banking. Why is it so hard?” Financial services, 2014.
3. Ting Yang Koh, Jermy Prenio, “Managing cloud risk – some considerations for the oversight of critical cloud service providers in the financial sector,” Bank for International Settlements, 2023.
4. Pańkowska M., “Knowledge worker tasks in system architecture,” Business Informatics 1(51), 2019.
5. Gellweiler C., “IT architects and IT-business alignment: a theoretical review,” Procedia Computer Science 196, pp. 13-20, 2022.
6. Harrison W., “A flexible method for maintaining software metrics data: a universal metrics repository,” Journal of Systems and Software, vol. 72, issue 2, pp. 225-234, July 2004.
7. Blackburn C., Valerdi R., “Navigating the Metrics Landscape: An Introductory Literature Guide to Metric Selection, Implementation, & Decision Making,” 7th Annual Conference on Systems Engineering Research (CSER), 2009.
8. Hossin M., Sulaiman M.N., “A review on evaluation metrics for data classification evaluations,” International Journal of Data Mining & Knowledge Management Process (IJDKP), vol.5, no.2, March 2015.