International Journal of Advanced Culture Technology Vol.9 No.4 345-354 (2021)
DOI https://doi.org/10.17703/IJACT.2021.9.4.345

IJACT 21-12-41

More Than 40 Percent of Data Unnecessarily Redundant in Corporate Databases

Jenghyearn Moon

Henley Business School, University of Reading
j.moon@pgr.reading.ac.uk

Abstract

Data quality issue in information systems is analyzed with focus on conceptual data modeling. Extensive investigation through triangulation of case studies is attempted to find how much extent inappropriate data modeling practices exercised in real workplace environment. It is revealed that more than 40 percent of data adversely contributed to unnecessary data redundancy, i.e., the level of data obesity is over 40 percent. Another contribution of this paper lies in excavation of all the categories of inappropriate data modeling practices, which has been previously only partially uncovered in the literature. New findings in this paper prove that the extent of inappropriate modeling is more serious that previously reported.

Keywords: *Conceptual Data Modeling, Information Systems, Database, Unnecessary Data Redundancy, Data Obesity*

1. INTRODUCTION

1.1 Difficulty of Achieving Quality Database Design

Designing a database is not a deterministic task in that two database designers may follow all the same strict, rigid rules on normalization principles for a given problem, but often they unfortunately happened to generate different data design layouts. This is inherent to the creative and even artistic nature of database design. Therefore, it is true that there is no exact correct answer to this dilemma in database design. The important thing consequently is that which one is appropriately designed in terms of design quality criteria for data. Nevertheless, there is an design technique that can be applied to every circumstance, and to follow them is the best way to get to a database that performs at its best. It may seem that data modeling is a lot of work without guaranteed benefits, since it involves the complexity of a system to be modeled, personalities of modelers, and even the culture of organization as in any other area of modeling software project management [1]. This paper investigates the issue of data modeling in the in the context of business environment with a focus on the quality of conceptual data modeling, with technological analysis. The significance of data modeling lies in process-independent representation of business data.

The value of data modeling has always been to establish the 'single version of the truth' of one's business or to have a 'single 360 degree view of the business' [2]. However, there is an opinion divide on whether data modeling is an art or science even in the database community. Some say it should be a science [3] as there are

Manuscript received: November 2, 2021 / revised: November 15, 2021 / accepted: December 3, 2021
Corresponding Author: j.moon@pgr.reading.ac.uk
Tel: +44(0)-118-378-4418
PhD Student, Henley School of Business, University of Reading | henley.ac.uk/bisa

many rules to comply with, but others say that it is an art rather than science as even rules are interpreted differently to a significant extent depending upon the designers themselves. Overall, it can be viewed not only science as there are obviously certain technical skillset for it but also art as there are obviously behavioral skills for it in order to draw out successful outcome of data modeling. Nevertheless, there are common ground of understanding on how inappropriate is bad practices in data modeling. The most common concerns are to see if there is increase in chance of errors, their negative impact on future performance and reduction in portability and maintainability [3].

1.2 Data Modeling as Core Element of Information Systems

Data modeling constitutes a core element of business information systems as the success or failure in terms of data quality undermines the performance of information systems [4, 5]. However, there has been a constant concern with regard to data quality due to poor data modeling [6-8]. This concern is centered on particularly at the level of conceptual data modeling, since conceptual schema [9, 10] should be the center of development of information base as the formal specification of all the functional information system requirements [11]. In the professional information systems community, two of the currently most popular development approaches have been the agile method [12] and the unified process [13]. Progress from conventional approach toward rapid applications development turned out to be enormous and impressive. Nevertheless, precise formulation of conceptual data modeling plays a significant part in neither of them, and as a matter of fact, unfortunately, the needs of conceptual data modeling in information systems development has been overlooked even with prevalence of agile methods and unified modeling.

The reason behind this neglect is that there has been a continuous lack of maturity in standard for information systems development [14]. In order to perform functions required for information system, the information system must possess a certain general knowledge about its domain and the functions it has to perform. Every such knowledge resides within a territory of conceptual data model, thus every information system needs to embody conceptual data model in it [15, 16]. Therefore, an information system is destined to fail to perform any useful function if it is without a conceptual data model [17]. It is obvious that there is an undisputed consent that, as opposed to most mature scientific disciplines, the discipline of information systems development does not exhibit a sound and widely accepted foundation of basic concepts. Numerous and diverse views abound in this field of study.

In general, this is caused by the lack of formal foundation [18], although there have been some hundreds of information systems development methods proliferated [19]. The resulting situation has often been referred to as the 'methodology jungle' [18, 20]. The attempts for formalism have been made but they failed, since they turned out too formal to be adopted and used by industry practitioners [21, 22]. This has unfortunately led to fuzzy and artificial concepts in information systems development methods as well as in conceptual data modeling, and as result most organizations and research groups have defined their own data modeling methods [23].

1.3 Objectives

The aim of paper is to analyze the data quality issue in business information systems and investigate its current practices in real workplace in order to explore their limitations and problems uncovered. To achieve that aim, the objective of paper is to investigate on the level of data quality in conceptual data modeling with measurement, not only qualitatively but also quantitatively, on basis of various metrics. The major metric will be the degree of unnecessary data redundancy in conceptual data model. The key research objectives are as follows. (1) The state-of-the-art data quality literature is investigated to explore if there is a satisfactory closure of answer to the questions regarding the conventional structured databases at the current wave of advancement is moving toward knowledge-based systems and big data systems. (2) In order to investigate data consistency issue at the level of conceptual data modeling, the entity-relationship modeling, rather than other semantic models, is explored to find out inappropriate cases of data modeling.

2. LITERATURE REVIEW

2.1 Conceptual Data Model as Evaluation Criterion of Data Quality

Conceptual data modeling has been valued as an important means to achieve data independence in a way of separation of upper logical data level from lower physical data level. In the literature, there are some dozens of semantic data modeling techniques proposed from late 1970s to late 1980s [28]. Entity-relationship (ER) modeling is one of the representative modeling technique since the advent of original version of such modeling proposed by Peter Chen [29]. Surveys in a form of interview to real world practitioners and experiments in learning data modeling technique by students revealed a consistent trend that they prefer ER model or extended versions of ER model than any other types of semantic data model [30]. The major reason for the preference toward ER models is that the E-R model has achieved a good equilibrium between expressive power on one hand and simplicity and generality in applicability on the other in accommodating the requirements of modern applications. While this model, like any other conceptual models, is limited on its expressive power, this limitation is a necessary trade-off for its formality, simplicity and wide applicability [30]. The other reasons include easiness in learning [31] and possibility of applying normalization theory at the higher conceptual level of abstraction. The idea is that the generation of normailzed relations is guaranteed at the conceptual model level [32, 33] by incorporating the notion of functional dependencies [27] as early as possible at the level of conceptual design.

2.2 Timeliness versus Consistency

2.2.1 Importance of Timeliness

According to recent discovery in pitfall of conceptual data modeling, even in the case of ER modeling [24], timeliness can be rather guaranteed through a rigid application of fundamental data model design principles so that data model design outcomes strictly in full compliance with database normalized forms straightforwardly. This may be in contrast to the conventional way of thinking that rigidity of relational model, per se, is a presumed cause to the impediment to timeliness.

2.2.2 Data Consistency Neglect in Ontology-Based Semantic Models

There might be a speculation that ontology-based models, such as RDF or OWL including UML [34], may contribute to ensure data consistency through use of thesauri. The integrity constraints used in relational database are for validating the integrity on the value of data, but integrity constraints used in ontology-based data models are very different to those in relational database as they are not interested in data value, instead by focusing on axioms, for instance, “every Event must have a Location where it takes place” [34]. It is evident that ontology-based approach started with presumption that data consistency issue for structured data has already been completely resolves, nevertheless, unfortunately, such presumption is not valid as the closure on many questions regarding that category of data consistency is still has not come even up to now. The contribution of ontology-based models is significant in terms of its ability in representing information architecture that are essential for the description of webpages, but due to their intrinsic nature of regarding data value consistency less important and their inherent ambiguity in distinguishing entities from relationships it is needed to exclude them from investigation on the issue of data consistency.

2.3 Pitfalls in Data Modeling

2.3.1 Illegitimate Use of Foreign Keys

One of the major source of poor data modeling lies in an inappropriate use of foreign keys [7, 24]. The use of a primary key as a foreign key is very strict and rigid in relational database design. A foreign key needs to exist only between an entity and its associated relationships directly adjacent to that entity. If some primary key of an entity appears in some distant entities or relationships as foreign key, i.e. not directly adjacent to that entity, it is out of normalcy, and this amounts to misuse of foreign keys. Note that there is a natural and automatic redundancy between primary key attribute(s) of entity and its directly adjacent foreign key attribute(s) of relationship, and this redundancy is called inevitable when conceptual data model is mapped into relational logical data model as the two tables, corresponding to entity and relationship contain the same attribute(s). Any other forms of use of attribute(s) as foreign key other than uses in this directly adjacent manner trigger data redundancy, which obviously becomes unnecessary.

Another main cause is the use of foreign keys in wrong places of entities or relationships, which can intensify the degree of data redundancy in a way that the same piece of data exists across multiple places [32]. It is reported that about half of data model design outcomes in practice have foreign key problems including dangling foreign key, which points to a primary key that is not there [24]. Note that it can make the data corrupted, nonsensical or inconsistent. Enforcement of referential integrity between primary keys and foreign keys is strictly limited to be guaranteed by database management systems only for their legitimate use [33]. DBMS is not responsible for maintaining data consistency for any other cases of poor use of foreign key, in which it is not partnered with its directly adjacent primary key [25]. Misuses in key propagation of foreign keys bring unnecessary data replication [7]. This case of pitfall will be investigated with a real-world example case in detail in Section 4.1.

2.3.2 Precarious Replication of Non-key Attributes

Other category of poor data modeling is redundancy accrued from careless replications of non-key attributes in wrong places [25]. The impacts of duplicates in databases, particularly in biological area, in practice were reported to be serious in that duplicates lead to data redundancies, which violate the principles of data normalization, and duplicates lead to inconsistencies and may have propagated impacts even after being detected.

3. ANALYSIS OF INDUSTRY CASE STUDIES

3.1 Analysis on Misuse in Foreign Keys in Conceptual Data Models

Note that Figure 1 is originally from Rhee's work [25] but it is redrawn here, rather than reprinted, in this paper with emphasis for new findings of erroneous data modeling, which were uncovered by the author of this paper apart from Rhee's discovery [25], with highlights in red and blue. If we scrutinize Figure 1 carefully, the case of some primary key of an entity appears in some entities or relationships as foreign key, which are distant, i.e., not directly adjacent, from that entity can be found in many places of entities or relationships. One of such case is an attribute 'Product_Code' in the entity 'Products_Displayed' in the middle centre area of Figure 1 as it cannot have 'Product_Code' as its foreign key. The origin of 'Product_Code' is from an entity

‘Products_MasterFile’, situated below ‘Product_Displayed’ in the middle of the figure, and that attribute is designated as the primary key of that entity. If we look at carefully, the only data elements which are directly connected to the entity ‘Products_MasterFile’ are the five relationships, which are ‘Products_Advertised’, ‘Products_Bundled_Components’, ‘Products_Shopped’, ‘Inventory’ and ‘Products_Ordered’, all of these are eligible to contain the primary key of ‘Products_MasterFile’ as their respective foreign key. However, ‘Products_Displayed’ obviously contains ‘Product-Code’ as its attribute, moreover as one of its foreign keys, which clearly violates the design rule in ER model twice. Use of attribute(s) as foreign key other than this manner of directly adjacent places trigger data redundancy, which becomes unnecessary. Note that ‘Product_Code’ in ‘Product_Displayed’ is not only incorrect but also causes this manner of unnecessary redundancy as there is no ground for that attribute extant there. In other words, there is no need at all for it to be included there.

3.2 Analysis on Incompliance to Legitimate Data Modeling

It was reported that real-workplace practitioners actually used non-normal forms rather than normal forms either unwittingly or illiterately under the assumption that their applications of denormalization would let the design outcome remain within the boundary of normal forms [32]. It was demonstrated that the design outcome of ‘E-Commerce application’ in Rhee’s work [32] exhibits a high ratio of unnecessary data replication, surprisingly 40 percent of unnecessary redundancy. In this paper, it was uncovered that there are three more new malpractices of flaw other than what were found in [32] for the E-commerce application as follows, from Section 4.3 to Section 4.5.

3.3 Incongruence against Legitimate Data Modeling Constructs

In the example of E-Commerce application in Rhee’s work [32], 44 cases of unnecessary data redundancy were found from the Figure 1 in [32] (in p. 48) and all of them are shown in the Figure 2 in [32] (in p.50). If all of 44 cases are removed from that Figure 2 [32], the resulting ER model comes out as Figure 2, with 29 components of entity and relationship as below. It turned out to be that many parts of data constructs, 23 cases altogether, do not comply with ‘E-R-E’ construct, meaning that Entity followed by Relationship followed by Entity. To show what those 23 cases are clearly, all of them are numbered in red color in Figure 2. Since there are 29 data entities in total in Figure 2, this implies that 76 percent, 23 out of 29, of the entire design is incongruent to data construct legitimacy. To measure exactly how much extent the conceptual data model in Figure 2 is deviated from or violated legitimacy, there turns out to be data modeling works end up simply with ‘E-E’, or ‘’E-R’ or ‘R-R’, which respectively indicates that either relationship, entity or entities are intentionally missing. There are 11 cases of ‘E-R’, 5 cases of ‘R-R’, and 4 cases of ‘E-E’ in Figure 2. The cases of ‘E-E’ and ‘E-R’ implies that either relationship or entity conspicuously subsumes in a component directly adjacent to it. Note that glancing at the overall framework picture of data constructs is important rather than legibility of individual attribute.

3.4 Misuse of Non-key Attributes

Another new finding in this paper that is not dealt with in the E-Commerce application [32] is the cases of misuse of non-key attributes in a way that unnecessary data redundancy is exacerbated. The attribute ‘Point_Awarded’ is one of them. That attribute was used in 9 different data components: 6 relationships of ‘Products_Shopped’, ‘Siding_Price’, ‘MemberGradeTable’, ‘SpecialPrice’, ‘Products_Ordered’ and ‘MemberInfo’, and 3 entities of ‘Order’, ‘Products_Displayed’ and ‘Point_AccumulatedLog’ (Note that all such 9 non-key attributes denoting 'Point_Awarded' are printed in blue in Figure 1). However, in reality the attribute 'Point_Awarded' is something that should appear in the relationship 'Payment/Collection' as one of its non-key attributes. Therefore, instead of appearing in 'Payment/Collection' in a sensible manner, it appeared elsewhere nine times unnecessarily, consequently contributing to aggravate the entire degree of unnecessary data redundancy far more than 40 percent, which is the exact figure discovered as the ratio of unnecessary data redundancy in [32]. If these 9 cases are added to the cases of unnecessary data redundancy uncovered in [32], which was 150 out of 280, the degree of unnecessary data redundancy will be increased, since 159 out of 280 turns out to be 0.57, which means 57 percent of data are redundant, amongst which 42 percent of data are unnecessary redundant. Note that 15 percent of inevitable data redundancy, revealed in [32], need to be removed from 57 percent, consequently the ratio of unnecessary data redundancy yields to be 42 percent. This vindicates that the ratio of unnecessary data redundancy is not 40 percent or less than 40 percent of corporate database but it actually exceeds 40 percent. Note that replication of any attribute, regardless of whether it is key attribute or non-key attribute, is classified as unnecessary. The only exception to this is the use of foreign key in a legitimate way, thus in this regard any replication of non-key attributes becomes unnecessary. This implies that more than 40 percent of data in real workplace corporate database are unnecessary redundant.

4. RESULT AND DISCUSSION

4.1 Interpretation of Result

4.1.1 Implications of New Findings in This Paper

In addition to the new findings with regard to the degradation toward non-first normal form, there appears one more irony in the E-commerce application in the total number of data construct components with denormalization in exercise and that after the complete removal of all the unnecessary redundancies. Refer to Figure 1 for the original data model, and Figure 3 below for the data model with unnecessary data redundancy completely removed (Reprinted here for convenience for readers). Note that the latter case is fully compliant to the 3NF design, but it does not make sense to find that the former, with denormalization, is 40 whilst the latter, with normalizaiton, is 29. This abnormal situation proves that the denormalizations were practiced illegitimately in Figure 1.

The benefits collected after denormalization are conventionally perceived to include a less number of tables than a certain optimal design, for instance 3NF design, so that the total number of joins between tables required to get the answers for database queries might be minimized for boost in speed. Nonetheless, the irony of contradictory reversal in the number of resultant data tables the two cases above invalidates the conventional perception on advantage of reduced number of joins owing to denormalization. Why does this happen in the case of E-Commerce application? The answer is the unnecessary data redundancy caused by misuse of foreign keys and misuse of non-key attributes, which adversely affects to draw out the poor database design after all.

4.1.2 Use of a New Notion of Data Obesity in Lieu of Normal Forms

Understanding on normal forms seems still not be easy for database design novices and even to database design experts. The major reason behind this difficulty is partly because of none of formulaic approach to data modeling available and partly because of the relational database firmly based on set theory, which many people have a hard time in fully understanding the logic behind profound mathematical theories in studying. The normal forms are useful in optimising the data redundancy, and given that the ratio of unnecessary data redundancy, i.e. data obesity, can be controlled [32], a new approach of controlling the optimal level of data redundancy based on data obesity seems to be feasible rather than sticking with the conventional normal form-based approach.

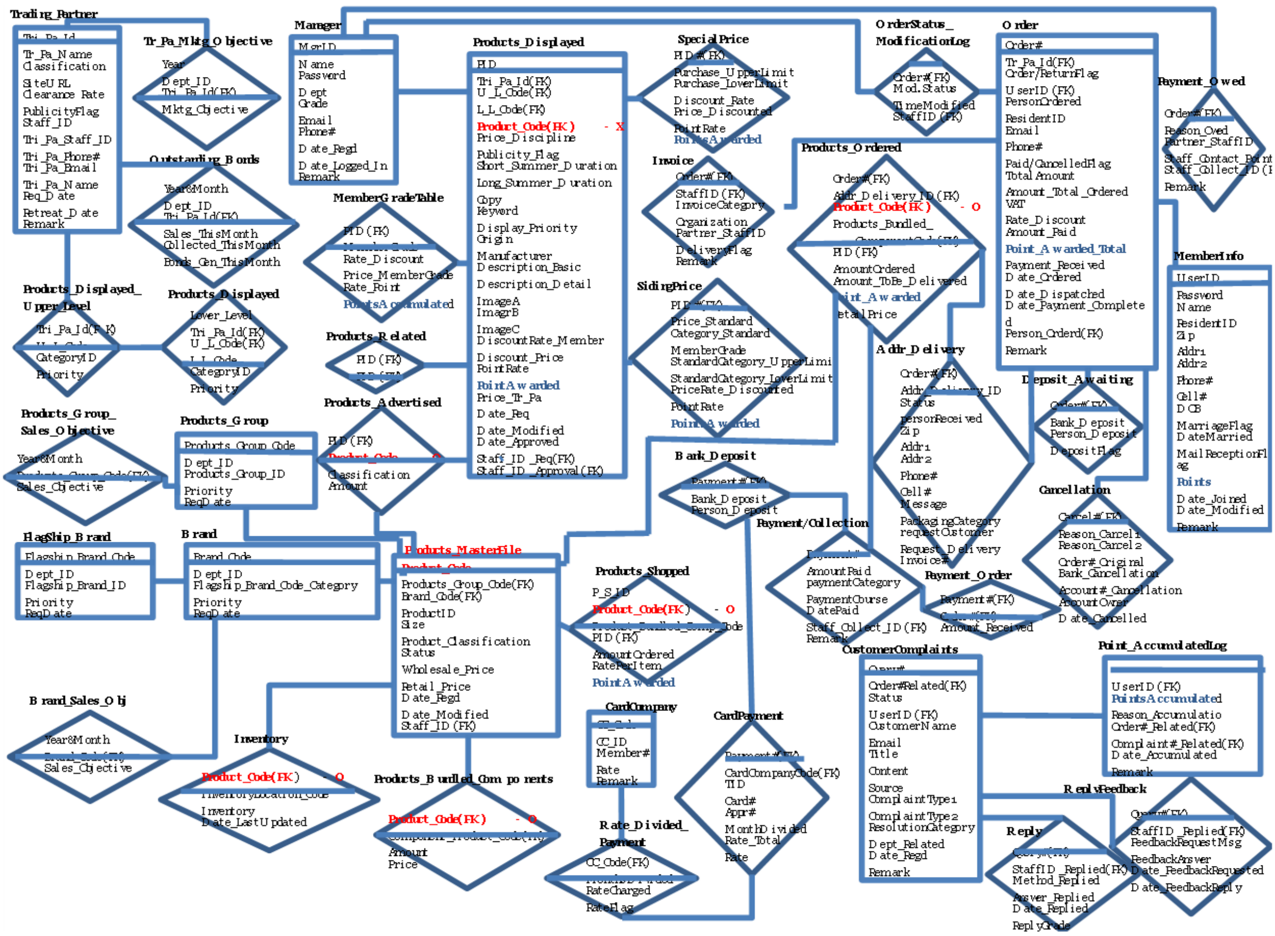

Figure 1. Example Conceptual Data Model for E-Commerce Application; Redrawn with Highlights

4.2 Ambiguity Embedded in the Process of Denormalization Practices

Although there were many studies on loss of data quality in terms of proliferation of data redundancy in the literature [5-7, 34], many of them are failed or limited in distinguishing between inevitable data redundancy and unnecessary data redundancy at the microscopic level of data attribute. The findings on unnecessary data redundancy due to precarious denormalization echo in the heuristic observation of some 40 percent of attributes that are unnecessarily redundant in real workplace databases [32] and even to some 60 percent of unnecessary redundancy. The other category of poor data modeling is redundancy accrued from careless replications of non-key attributes, particularly, throughout the course of denormalization [25]. The process of normalization is quite straightforward and there is no way to become biased or deviated from this strict course of normalization from 1NF to BCNF. However, the process of denormalization is not straightforward as there is no discipline on how far normal form is retreated. Consequently, causing too much freedom in exercising denormalization process at the hands of human data modelers. It was reported that there are even many cases in which database schema do not comply even with 1NF, which is the lowest level of normal forms, despite denormalizations are attempted and thought to be successful to practitioner's judgement [25-27].

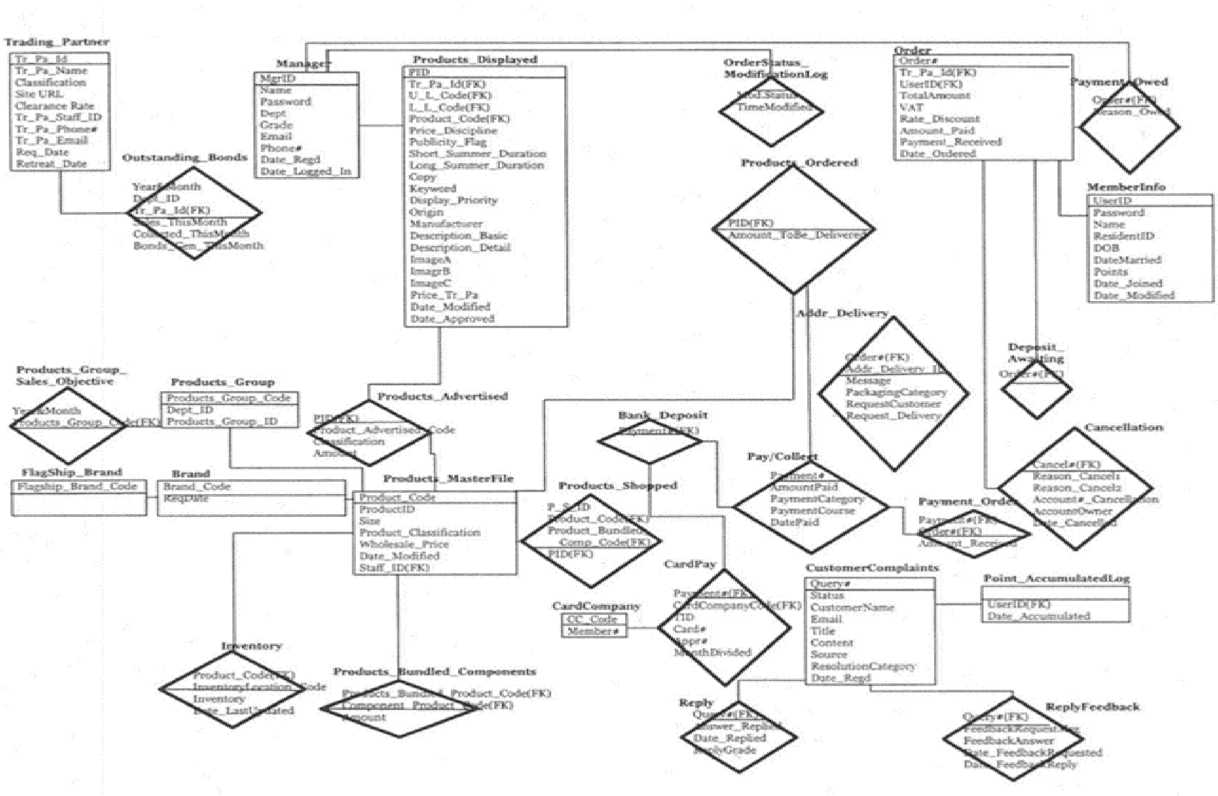

Figure 2. Cases of Anomaly of Not Abiding to ‘E-R-E’ Modeling Construct

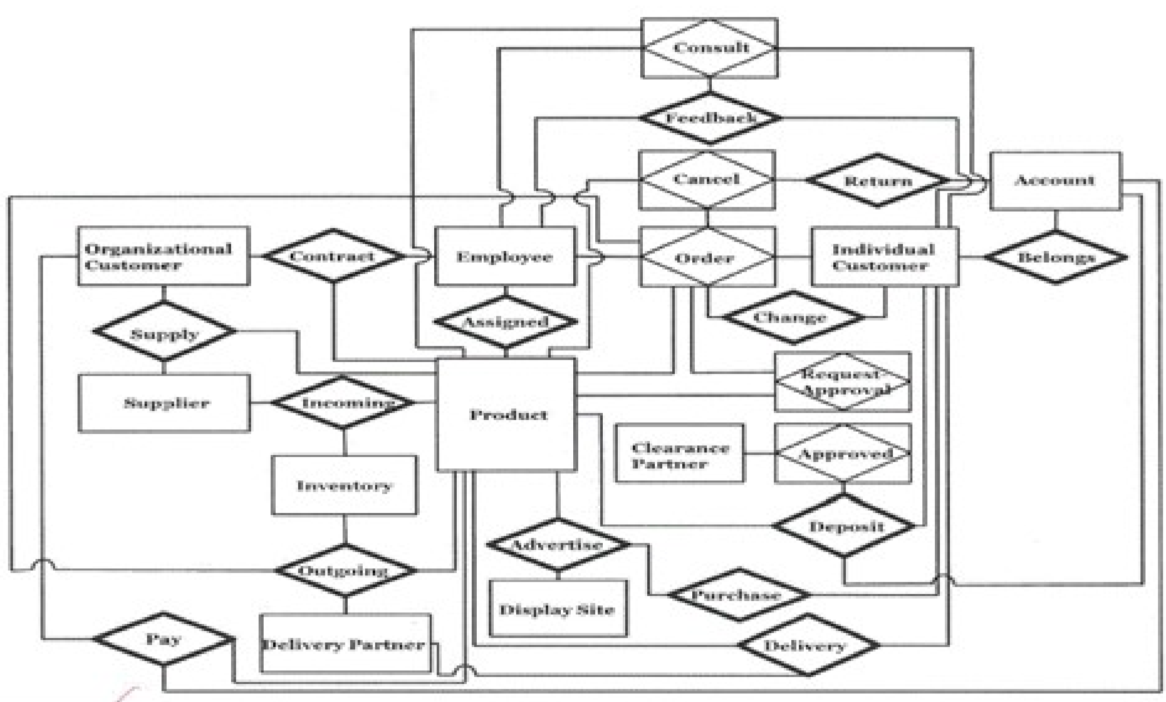

Figure 3. E-Commerce Model with Unnecessary Data Redundancy Removed

5. CONCLUSION

In this paper, the issue of data quality for conceptual data modeling in business information systems is analyzed. The major contribution of this paper lies in revealing the degree of unnecessary data redundancy through case studies in real workplace. In conclusion, this paper asserts that the ratio of unnecessary data redundancy in databases in reality is over 40 percent of data stored in them. This study lacks a solid foundation in resolution of data obesity reduction as it focused on evidences of denormalization malpractices. Practical implication of what the findings in this paper lies in a possibility that such reduction scheme can be devised to substitute the traditional normalization technique. Thus, it is needed to develop such scheme in a rigorous way in the future.

REFERENCES

[1] M. Thomsett, *The Little Black Book of Project Management*, American Management Association. 2010, New York, NY: AMACOM, 272 pages, 2010.
[2] L. Moss and S. Hoberman, *The Importance of Data Modeling as a Foundation for Business Insight*, Data Modeling and Business Insight, Realized Design, April 2008.
[3] H. Konelis, *Data Modeling: art or Science?*, SQLServerFast. 2008.
[4] J. Harris and S. Hoberman, *Data Modeling Made Simple with Erwin Data Modeler*. Technics Publications, NJ. U.S.A., 538 pages, 2020.
[5] F. Montans, F. Chinesta, R. Gomez-Bombarelli and J. Kutz, *Data-Driven Modeling and Learning in Science and Engineering*, Data-Based Engineering Science and Technology, Vol. 347, pp. 845-855, 2019.
[6] A. Haug, F. Zachariassen and D. van Liempd, *The Cost of Poor Data Quality. Journal of Industrial Engineering and Management*, Vol. 4, No. 2, pp. 168-193, 2011.
[7] C. Mancas, *Conceptual Data Modeling and Database Design: A Fully Algorithmic Approach*. Volume 1, Apple Academic Press, 698 pages, 2021.
[8] G. Sanders and S. Shin, *Denormalization Effects on Performance of RDBMS*, In Procs. 34th Hawaii International Conference on System Sciences, Vol. 3, pp. 1-9, 2001.
[9] A. Olive, *Conceptual Schema-Centric Development: A Grand Challenge for Information System Research*. In Procs. 17th International Conference on CAiSE, 13-17 June, In O. Pastor and J. Falcao e Cunha, (eds), LNCS, Vol. 3520, pp. 1-15, 2005.
[10] A. Tort and A. Olive, *An Approach to Testing Conceptual Schemas. Data & knowledge Engineering*, Vol. 69, pp. 598-618, 2010.
[11] A. Fayoumi and P. Loucopoulos, *Conceptual Modeling for the Design of Intelligent and Emergent Information Systems. Expert Systems with Applications*, Vol. 59, pp. 174-194, 2016.
[12] K. Beck, *Extreme Programming Explained: Embrace Change*, 2nd (ed), Boston, USA, 224 pages, 2005, Addison-Wesley.
[13] I. Jacobson, G. Booch and G. Rumbaugh, *The Unified Software Development Process*, Addison-Wesley, 463 pages, 1999.
[14] C. Rich and R. Water, *Automatic Programming: Myths and Prospects*, IEEE Computer, Vol. 21, No. 8, pp. 40-51, 1998.
[15] R. May, *Forging a Silver Bullet from the Essence of Software*, IBM Systems Journal, Vol. 33, No. 1, pp. 20-45. 1994.
[16] J. Sowa, *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks Cole Publishing. 594 pages, 2000.
[17] J. Mylopoulos, *Representing Software Engineering Knowledge. Automated Software Engineering*, Vol. 4, pp. 291–317. Kluwer Academic., 1997.

[18] A. Hofstede and T. Weide, *Formalisation of Techniques: Chopping down the Methodology Jungle. Information and Software Technology*, Vol. 34, No. 1, pp. 57-65, 1992.

[19] F. Stolterman, B. Fitzgerald and N. Russo, *Information Systems Development - Methods-in-Action*, McGraw-Hill, 2002.

[20] D. Avison and G. Fitzgerald, *Methodologies for Developing Information Systems: A Historical Perspective*, in Procs. IFIP 19th World Computer Congress on Past and Future of Information Systems: 1976–2006 and Beyond: Information System Stream, August 21–23, Santiago, Chile, 27-38, 2006.

[21] F. Baader, D. Calvanese, D. Mcguinness and D. Nardi, *The Description Logic Handbook: Theory, Implementation, and Applications*, 2nd ed, Cambridge University Press, 510 pages, 2007.

[22] C. Rolland and N. Pratkash, *From Conceptual Modeling to Requirement Engineering*, Annals of Software Engineering, Vol. 1, pp. 151-176, 2000.

[23] E. Safan, R. Meredith and F. Burstein, *Towards a Business Intelligence Systems Development Methodology: Drawing on Decision Support and Executive Information Systems*, in Procs. Pacific Asia Conference on Information Systems, Association for Information Systems Library, 2016.

[24] T. Nagle, T. Redman, T and D. Sammon, *Waking Up to Data Quality. The European Business Review*, 12 May 2018.

[25] H. Rhee, *Corporate Data Obesity: 50 Percent Redundant*, Global Journal of Computer Science and Technology, Vol. 10, No. 5, pp. 7-11. 2010.

[26] F. Martinez, *Bad Practices in Database Design: Are You Making These Mistakes?*, Developers, 2021.

[27] W. Lemahieu, S. Broucke and B. Baesens, *Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Small and Big Data*. Cambridge University Press, 1807 pages, 2018.

[28] R. Hull and R. King, *Semantic Database Modeling: Survey, Applications, and Research Issues*, ACM Computing Surveys, Vol. 19, No. 3, pp. 201-260, 1987.

[29] P. Chen, *The Entity-Relationship Model - Toward a Unified View of Data*, ACM Transactions on Database Systems, Vol. 1, No. 1, pp. 9-36, 1976.

[30] A. Badia, *Entity-Relationship Modeling Revisited*, SIGMOD Record, Vol. 33, No. 1, pp. 77-82, 2004.

[31] S. Jarvenpaa and J. Machesky, *Data Analysis and Learning: An Experimental Study of Data Modeling Tools*, International Journal of Man-Machine Studies, Vol. 31, pp. 367-391, 1989.

[32] H. Rhee, *State-of-The-Art Worldwide Widespread ERP-borne Misuse of Data*, International Journal of Innovative Trends in Engineering, Vol. 37, No. 1, pp. 47-53, 2018.

[33] C. Ordonez and J. Garcia-Garcia, *Referential Integrity Quality Metrics. Decision Support Systems*, Vol. 44, pp. 495-508, 2008.

[34] D. Allemang and J. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, Burlington, MA: Morgan Kaufmann, pages 384, 2011.