

BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

Council of Doctoral University Studies

Doctoral School

Economic Informatics

**AGILE BUSINESS INTELLIGENCE SOLUTIONS FOR
INTEGRATING LARGE VOLUMES OF DATA**

ANDREEA VINEȘ

PhD Coordinator: Prof. Univ. Dr. Ana Ramona BOLOGA

Bucharest, 2025

CONTENT

SUMMARY(RO)	i
SUMMARY	ii
ACKNOWLEDGMENTS	iii
1. INTRODUCTION	4
1.1. Research topic and motivation	4
1.2. Objectives of the research program	5
1.3. Hypotheses and research methodology	7
1.4. Structure of the paper	10
2. CURRENT STATE OF RESEARCH	13
2.1. Introduction to Business Intelligence	14
2.1.1. Definition of Business Intelligence and its evolution	14
2.1.2. Business Intelligence in decision-making and organizational analysis	17
2.2. Introduction to Big Data	18
2.2.1. Characteristics of Big Data	19
2.3. Architectures for data management	20
2.3.1. OLTP vs. OLAP architecture	21
2.3.2. Data Warehouses and Data Lakes	21
2.3.3. Modern architectures: Data Mesh and Data Fabric	24

2.3.4. Data orchestration and processing models (ETL vs. ELT)	25
2.4. Artificial Intelligence in Business Intelligence	28
2.4.1. AI in Business Intelligence solutions	28
2.4.2. Large Language Models in Business Intelligence solutions	29
2.5. Personal contributions	31
3. TECHNOLOGIES USED IN DATA SOLUTIONS	32
3.1. Overview of on-premise and Cloud technologies for data solutions	34
3.2. Analysis of technologies used in data solutions – Case Study	37
3.2.1. Questionnaire presentation	38
3.2.2. Respondent analysis	39
3.2.3. Interpretation of results	42
3.3. Data Extraction techniques – Case Study	46
3.3.1. Web Scraping – Data extraction method	46
3.3.2. Automatic data extraction model	48
3.3.3. Web scraping challenges	52
3.4. Personal contributions	53
4. BUSINESS INTELLIGENCE ARCHITECTURES AND DATA MODELS	55
4.1. Dimensional modeling in Business Intelligence architecture	56
4.2. Data Vault Architecture – Concept and Implementation	60
4.3. Analysis of data architectures used in companies – Case Study	63

4.3.1. Interview structure	63
4.3.2. Interview results	65
4.3.3. Conclusions and emerging trends	67
4.4. Comparative analysis of Dimensional Modeling and Data Vault – Case Study	68
4.4.1. Implementation process stages	69
4.4.2. Dataset description	71
4.4.3. Model design	72
4.4.4. Results and conclusions	75
4.5. Personal contributions	79
5. AUTOMATED GENERATION AND TESTING OF DATA MODELS	82
5.1. Objectives of the proposed architecture	85
5.2. Solution architecture for automated Data Vault model generation	87
5.2.1. Need for automation in data modeling processes	87
5.2.2. Strategies for exploring Large Language Models in BI	89
5.2.3. General processing workflow	91
5.2.4. Determining data input	94
5.2.5. Data Vault model generation	95
5.2.6. Model validity evaluation	97
5.2.7. Model correction	98
5.2.8. Saving model in database	100

5.2.9. Source-to-Destination mapping generation	101
5.2.10. Saving and exporting Source-to-Destination mapping	102
5.3. Methodology for model validity testing	102
5.3.1. Data model quality	103
5.3.2. Structural integrity component	107
5.3.3. Data Vault standards validation component	111
5.3.4. Performance and scalability component	115
5.3.5. Validity coefficient determination	117
5.3.6. Conclusions and challenges	118
5.4. Personal contributions	120
6. IMPLEMENTATION AND TESTING OF AGILE ARCHITECTURE BASED ON LARGE LANGUAGE MODELS	122
6.1. Implementation of Business Intelligence architecture	123
6.1.1. User Interface	124
6.1.2. Data model generation and processing module	127
6.1.3. Model validity testing and management module	130
6.1.4. Source-to-Destination mapping generation	143
6.2. Testing the architecture on various datasets	146
6.2.1. Description of the datasets	147
6.2.2. Selection of Large Language Models	152

6.2.3. Testing process and validity coefficient generation	153
6.2.4. Metrics analyzed and test results	154
6.3. Conclusions and future directions	173
6.4. Personal contributions	174
7. CONCLUSIONS	175
7.1. Final conclusions	175
7.2. Personal contributions	177
7.3. Future research directions	179
BIBLIOGRAPHY	181
APPENDICES	192
1. Glossary of terms and abbreviations	192
2. List of figures	195
3. List of tables	197
4. List of algorithms	198
5. List of published papers	199

SUMMARY

The present paper begins with an in-depth analysis of the current state of research in the field of Business Intelligence, with a particular focus on the challenges encountered in the integration and modeling of large volumes of data. In the initial phase, data was collected using both qualitative and quantitative methods, such as surveys and interviews, with the goal of understanding current practices, actual industry needs, and expert perspectives on data solutions. Based on these insights, parallel implementations of Business Intelligence solutions were carried out to highlight the differences between various models and architectures used in practice.

Building upon the conclusions of this exploration phase, the research proposes agile architecture for data modeling and integration, centered around the use of Large Language Models. A validation model was developed to assess the quality and coherence of the data models generated by Large Language Models based on a clearly defined set of rules. The primary objective was to evaluate the extent to which the generated models comply with established standards and whether they can be effectively reused in the design stages of Business Intelligence solutions, while reducing manual and repetitive modeling tasks.

To test the applicability of the proposed architecture, a functional application prototype was developed, integrating Large Language Models into the data modeling process. This prototype was tested on various datasets and using different model versions, with a focus on robustness, flexibility, and ease of integration into existing BI workflows. The results support the potential of using Large Language Models as complementary tools within modern data architectures, offering an agile framework for the rapid and adaptive development of analytical solutions.

Keywords: *Business Intelligence, Agile Architecture, Data Modeling, Large Language Models, Automated Data Integration.*