

**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**

**Consiliul pentru Studii Universitare de Doctorat**

***Școala Doctorală***

**Informatică Economică**

**SOLUȚII AGILE DE BUSINESS INTELLIGENCE PENTRU  
INTEGRAREA VOLUMELOR MARI DE DATE**

**ANDREEA VINEȘ**

Conducător de doctorat: Prof. Univ. Dr. Ana Ramona BOLOGA

**București, 2025**

# CUPRINS

|                                                                              |     |
|------------------------------------------------------------------------------|-----|
| REZUMAT                                                                      | i   |
| SUMMARY                                                                      | ii  |
| MULȚUMIRI                                                                    | iii |
| 1. INTRODUCERE                                                               | 4   |
| 1.1. Tema de cercetare și motivația                                          | 4   |
| 1.2. Obiectivele programului de cercetare                                    | 5   |
| 1.3. Ipotezele și metodologia cercetării                                     | 7   |
| 1.4. Structura lucrării                                                      | 10  |
| 2. STADIUL ACTUAL AL CERCETĂRII                                              | 13  |
| 2.1. Introducere în Business Intelligence                                    | 14  |
| 2.1.1. Definirea Business Intelligence și evoluția sa                        | 14  |
| 2.1.2. Business Intelligence în luarea deciziilor și analiza organizațională | 17  |
| 2.2. Introducere în Big Data                                                 | 18  |
| 2.2.1. Caracteristici Big Data                                               | 19  |
| 2.3. Arhitecturi pentru gestionarea datelor                                  | 20  |
| 2.3.1. Arhitectura clasică OLTP vs. OLAP                                     | 21  |
| 2.3.2. Data Warehouses și Data Lakes                                         | 21  |
| 2.3.3. Arhitecturi moderne de tip Data Mesh și Data Fabric                   | 24  |
| 2.3.4. Modele de orchestrare și procesare a datelor (ETL vs. ELT)            | 25  |
| 2.4. Inteligența artificială în Business Intelligence                        | 28  |
| 2.4.1. Inteligența Artificială în soluții de Business Intelligence           | 28  |
| 2.4.2. Modelele lingvistice mari în soluții de Business Intelligence         | 29  |
| 2.5. Contribuții personale                                                   | 31  |

|        |                                                                                |    |
|--------|--------------------------------------------------------------------------------|----|
| 3.     | TEHNOLOGII UTILIZATE ÎN SOLUȚIILE DE DATE                                      | 32 |
| 3.1.   | Prezentarea tehnologiilor on premise și Cloud pentru soluțiile de date         | 34 |
| 3.2.   | Analiza tehnologiilor utilizate în soluțiile de date – Studiu de caz           | 37 |
| 3.2.1. | Prezentarea chestionarului                                                     | 38 |
| 3.2.2. | Analiza respondenților                                                         | 39 |
| 3.2.3. | Interpretarea rezultatelor                                                     | 42 |
| 3.3.   | Tehnici de extragere a datelor – Studiu de caz                                 | 46 |
| 3.3.1. | Web scraping – metodă de extragere a datelor                                   | 46 |
| 3.3.2. | Prezentarea modelului de extragere automată a datelor                          | 48 |
| 3.3.3. | Provocări în web scraping                                                      | 52 |
| 3.4.   | Contribuții personale                                                          | 53 |
| 4.     | ARHITECTURI DE BUSINESS INTELLIGENCE ȘI MODELE DE DATE                         | 55 |
| 4.1.   | Modelul Dimensional în arhitectura Business Intelligence                       | 56 |
| 4.2.   | Arhitectura Data Vault – concept și implementare                               | 60 |
| 4.3.   | Analiza arhitecturilor de date utilizate în cadrul companiilor – Studiu de caz | 63 |
| 4.3.1. | Structura interviului                                                          | 63 |
| 4.3.2. | Rezultatele interviului                                                        | 65 |
| 4.3.3. | Concluzii și tendințe emergente                                                | 67 |
| 4.4.   | Analiza comparativă între Modelarea Dimensională și Data Vault – Studiu de caz | 68 |
| 4.4.1. | Etapele procesului de implementare                                             | 69 |
| 4.4.2. | Descrierea setului de date                                                     | 71 |
| 4.4.3. | Proiectarea modelului                                                          | 72 |
| 4.4.4. | Rezultate și concluzii                                                         | 75 |
| 4.5.   | Contribuții personale                                                          | 79 |
| 5.     | GENERAREA ȘI TESTAREA AUTOMATĂ A MODELELOR DE DATE                             | 82 |

|                                                                                   |     |
|-----------------------------------------------------------------------------------|-----|
| 5.1. Obiectivele arhitecturii propuse                                             | 85  |
| 5.2. Arhitectura soluției pentru generarea automată a modelelor Data Vault        | 87  |
| 5.2.1. Necesitatea unui flux de automatizare a procesului de modelarea datelor    | 87  |
| 5.2.2. Necesitatea unui flux de automatizare a procesului de modelarea datelor    | 89  |
| 5.2.3. Fluxul general de procesare                                                | 91  |
| 5.2.4. Determinarea inputului de date                                             | 94  |
| 5.2.5. Generarea modelului Data Vault                                             | 95  |
| 5.2.6. Evaluarea validității modelului                                            | 97  |
| 5.2.7. Corecția modelului                                                         | 98  |
| 5.2.8. Salvarea modelului în baza de date                                         | 100 |
| 5.2.9. Generarea mapării sursă-destinație                                         | 101 |
| 5.2.10. Salvarea și exportarea mapării sursă-destinație                           | 102 |
| 5.3. Metodologie de testare a validității modelului                               | 102 |
| 5.3.1. Calitatea modelelor de date                                                | 103 |
| 5.3.2. Componenta integritatea structurală                                        | 107 |
| 5.3.3. Componenta de validare a standardelor Data Vault                           | 111 |
| 5.3.4. Componenta de performanță și scalabilitate                                 | 115 |
| 5.3.5. Determinarea coeficientului de validitate                                  | 117 |
| 5.3.6. Concluzii și provocări                                                     | 118 |
| 5.4. Contribuții personale                                                        | 120 |
| 6. IMPLEMENTAREA ȘI TESTAREA ARHITECTURII AGILE BAZATE PE MODELE LINGVISTICE MARI | 122 |
| 6.1. Implementarea arhitecturii de Business Intelligence                          | 123 |
| 6.1.1. Interfața cu utilizatorul                                                  | 124 |
| 6.1.2. Modulul de procesare și generare a modelelor de date                       | 127 |

|                                                                      |     |
|----------------------------------------------------------------------|-----|
| 6.1.3. Modulul de testare a validității și gestionare a modelelor    | 130 |
| 6.1.4. Generarea mapării de tipul sursă-destinație                   | 143 |
| 6.2. Testarea arhitecturii pe diverse seturi de date                 | 146 |
| 6.2.1. Descrierea seturilor de date utilizate                        | 147 |
| 6.2.2. Alegerea modelelor lingvistice mari                           | 152 |
| 6.2.3. Procesul de testare și generarea coeficienților de validitate | 153 |
| 6.2.4. Metrice analizate și rezultatele testării                     | 154 |
| 6.3. Concluzii și direcții viitoare                                  | 173 |
| 6.4. Contribuții personale                                           | 174 |
| 7. CONCLUZII                                                         | 175 |
| 7.1. Concluzii finale                                                | 175 |
| 7.2. Contribuții personale                                           | 177 |
| 7.3. Direcții viitoare de cercetare                                  | 179 |
| BIBLIOGRAFIE                                                         | 181 |
| ANEXE                                                                | 192 |
| 1. Glosar de termeni și abrevieri                                    | 192 |
| 2. Listă figuri                                                      | 195 |
| 3. Listă tabele                                                      | 197 |
| 4. Listă algoritmi                                                   | 198 |
| 5. Listă lucrări publicate                                           | 199 |

# REZUMAT

Lucrarea curentă pornește de la o analiză aprofundată a stadiului actual cercetării în domeniul Business Intelligence, punându-se accent pe provocările întâlnite în integrarea și modelarea volumelor mari de date. În prima etapă, au fost colectate date prin metode calitative și cantitative, precum chestionare și interviuri, cu scopul de a înțelege practicile actuale, nevoile reale din industrie și percepțiile specialiștilor cu privire la soluțiile de date. Pe baza acestor informații, s-au realizat implementări paralele de soluții de Business Intelligence, pentru a evidenția diferențele dintre diverse modele și arhitecturi utilizate în practică.

Pornind de la concluziile acestei faze exploratorii, cercetarea propune o arhitectură de tip „agil” pentru modelarea și integrarea datelor, ce este centrată pe utilizarea modelelor lingvistice mari. A fost elaborat un model de validare ce permite evaluarea calității și coerenței modelelor de date generate de LLM, pe baza unui set bine definit de reguli. Scopul principal a fost de a valida în ce măsură modelele generate respectă standardele impuse, și se pot utiliza ulterior în etapele de proiectare a soluțiilor de Business Intelligence, reducând sarcinile manuale și repetitive de realizare a modelelor.

Pentru a testa aplicabilitatea arhitecturii propuse, a fost realizat un prototip funcțional de aplicație ce integrează modele lingvistice mari în procesul de generare a modelelor de date. Acest prototip a fost testat pe diverse seturi de date și cu diferite variante de modele, urmărindu-se robustețea, flexibilitatea și ușurința de integrare în procesele de BI existente. Rezultatele obținute susțin potențialul utilizării acestor modele ca instrumente complementare în arhitecturile moderne de date, oferind un cadrul agil pentru dezvoltarea rapidă și adaptabilă a soluțiilor analitice.

**Cuvinte cheie:** *Business Intelligence, Arhitectură agilă, Modelare de date, Large Language Models, Integrare automată a datelor.*