

**BUCHAREST UNIVERSITY OF ECONOMIC STUDIES**  
Council for Doctoral Studies  
Economic Informatics Doctoral School

**DATA PROCESSING AND ANALYSIS IN THE CONTEXT OF  
IoT AND BIG DATA**

**Dragoş – Cătălin G. Barbu**

**PhD Supervisor: Prof. Univ. Dr. Adela Bâra**

Bucureşti, 2024

## Table of Content

<a href="#">Rezumat</a> .....	i
<a href="#">Multumiri</a> .....	ii
<b>1. INTRODUCERE</b> .....	<b>1</b>
<b>1.1. CONTEXT ȘI MOTIVAȚIE</b> .....	<b>1</b>
1.1.1. <a href="#">Context</a> .....	1
1.1.2. <a href="#">Motivare</a> .....	2
1.1.3. <a href="#">Metodologie</a> .....	3
1.1.4. <a href="#">Semnificația cercetării</a> .....	3
1.1.5. <a href="#">Structura tezei</a> .....	4
1.1.6. <a href="#">Obiectivele cercetării</a> .....	5
<b>1.2. SEMNIFICAȚIA STUDIULUI</b> .....	<b>6</b>
<b>2. STADIUL ACTUAL AL CERCETĂRII</b> .....	<b>8</b>
<b>2.1. BIG DATA: VIZIUNE GENERALĂ ȘI APLICAȚII</b> .....	<b>8</b>
2.1.1. <a href="#">Introducere</a> .....	8
2.1.2. <a href="#">Definiția și caracteristicile Big Data</a> .....	9
2.1.3. <a href="#">Tehnici și tehnologii de analiză a datelor mari</a> .....	13
<b>2.2. ANALIZA INFORMAȚIILOR DIGITALE DE NATURĂ FISCALĂ</b> .....	<b>13</b>
2.2.1. <a href="#">Cercetări și studii anterioare</a> .....	14
2.2.2. <a href="#">Big Data Analytics în context românesc</a> .....	16
2.2.3. <a href="#">Studii de caz în analiza datelor fiscale</a> .....	17
2.2.4. <a href="#">Caracteristicile datelor fiscale din casele de marcat electronice</a> .....	18
<b>2.3. CICLUL DE VIAȚĂ AL DATELOR</b> .....	<b>20</b>
<b>2.4. STUDIU ȘI PLANIFICARE ASUPRA INFORMAȚIILOR FISCALE DIN ECR-URI</b> .....	<b>23</b>
<b>2.5. COLECTAREA INFORMAȚIILOR DIGITALE DE NATURĂ FISCALĂ</b> .....	<b>24</b>
2.5.1. <a href="#">IOT – INTERNETUL OBIECTELOR (INTERNET OF THINGS)</a> .....	25
<b>2.6. PREPROCESAREA ȘI CURĂȚAREA DATELOR</b> .....	<b>28</b>
<b>2.7. ANALIZA DATELOR</b> .....	<b>28</b>
2.7.1. <a href="#">Analiză în timp real și flux de date</a> .....	28
2.7.2. <a href="#">Analiza exploratorie a datelor</a> .....	29
2.7.3. <a href="#">Tendințe în analiza informațiilor de tip Big Data. Tehnici și metode de extragere a datelor</a> .....	30
2.7.4. <a href="#">Modelul de regresie simplă</a> .....	31
2.7.5. <a href="#">Validarea modelului de regresie</a> .....	32
2.7.6. <a href="#">Machine learning</a> .....	34

2.7.7.	<a href="#">Paradigme ale învățării nesupravegheate</a>	35
2.7.8.	<a href="#">Clusterizare</a>	38
2.7.9.	<a href="#">Estimarea densității</a>	38
2.8.	<a href="#">VIZUALIZARE ȘI RAPORTARE</a>	38
2.9.	<a href="#">CONFIDENȚIALITATE ȘI PROVOCĂRI DE SECURITATE</a>	39
2.10.	<a href="#">CONFIDENȚIALITATE ȘI CONSIDERAȚII ETICE</a>	39
2.11.	<a href="#">IDENTIFICAREA LACUNELOR DE CERCETARE</a>	40
2.12.	<a href="#">REZUMAT ȘI CONTRIBUȚII PERSONALE</a>	41
3.	<a href="#">METODOLOGIA</a>	43
3.1.	<a href="#">ETAPELE ȘI OBIECTIVELE METODOLOGIEI</a>	43
3.2.	<a href="#">COLECTAREA DATELOR</a>	45
3.2.1.	<a href="#">Sursele datelor</a>	45
3.2.2.	<a href="#">Procedura de colectare</a>	46
3.2.3.	<a href="#">Considerații tehnice și de securitate</a>	49
3.3.	<a href="#">PRELUCRAREA ȘI ANALIZA DATELOR</a>	49
3.3.1.	<a href="#">Curățarea datelor</a>	49
3.3.2.	<a href="#">Instrumente și tehnici de analiză</a>	50
3.3.3.	<a href="#">Interpretarea rezultatelor</a>	52
3.4.	<a href="#">LIMITĂRILE METODOLOGIEI</a>	53
3.4.1.	<a href="#">Limitările legate de colectarea datelor</a>	53
3.4.2.	<a href="#">Limitările legate de prelucrarea datelor</a>	53
3.4.3.	<a href="#">Limitări legate de analiza datelor</a>	53
3.4.4.	<a href="#">Limitări legate de interpretarea și utilizarea datelor</a>	54
3.4.5.	<a href="#">Limitările contextului românesc</a>	54
3.5.	<a href="#">EVALUAREA ȘI VALIDAREA METODOLOGIEI</a>	54
3.5.1.	<a href="#">Validitatea datelor</a>	54
3.5.2.	<a href="#">Fiabilitatea datelor</a>	55
3.6.	<a href="#">REZUMAT ȘI CONTRIBUȚII PERSONALE</a>	57
4.	<a href="#">ARHITECTURA ȘI SPECIFICAȚIILE DE PROIECTARE ALE PLATFORMEI DE COLECTARE ȘI ANALIZĂ A DATELOR</a>	59
4.1.	<a href="#">TIPURILE DE INFORMAȚII</a>	60
4.2.	<a href="#">TEHNOLOGII SI ARHITECTURI</a>	63
4.3.	<a href="#">Colectarea de date</a>	65
4.3.1.	<a href="#">Transmiterea datelor</a>	65
4.3.2.	<a href="#">Transmiterea fișierelor XML</a>	67
4.4.	<a href="#">Preprocesarea datelor</a>	67

<a href="#">4.5. INTEGRAREA DATELOR</a>	68
<a href="#">4.6. VALIDAREA ȘI PREZENTAREA DATELOR</a>	68
<a href="#">4.7. ANALIZA DATELOR</a>	69
<a href="#">4.8. PUBLICAREA DATELOR</a>	69
<a href="#">4.9. REZUMAT ȘI CONTRIBUTII PERSONALE</a>	69
<a href="#">5. IMPLEMENTAREA PLATFORMEI</a>	71
<a href="#">5.1. MODUL PRIMIRE DATE</a>	71
<a href="#">5.2. MODUL PREPROCESARE DATE</a>	72
<a href="#">5.3. MODUL INTEGRAREA DATELOR</a>	74
<a href="#">5.4. MODUL PREPARARE/ PREGĂTIRE A DATELOR</a>	76
<a href="#">5.5. MODUL ANALIZA DATELOR (DATA ANALYTICS)</a>	77
<a href="#">5.6. MODULUL PUBLICARE REZULTATE</a>	82
<a href="#">5.7. REZUMAT ȘI CONTRIBUTII PERSONALE</a>	84
<a href="#">Comentariile</a>	86
<a href="#">Concluziile</a>	87
<a href="#">CONTRIBUȚII ORIGINALE</a>	89
<a href="#">LISTA LUCRĂRILOR ORIGINALE</a>	89
<a href="#">Referințele bibliografice</a>	92
<a href="#">Lista abrevierilor și acronimelor</a>	100
<a href="#">Lista figurilor/ graficelor</a>	100
<a href="#">Lista tabelor</a>	100
<a href="#">Lista anexelor</a>	101

**Keywords:** Big Data technics, Cloud computing, Data aggregation, Data enrichment, Digital fiscal data, Internet of Thing, Data mining, Data Storage, Big Data processing, Big Data Analytics, Machine learning, Electronic Cash Register

## **Summary:**

The doctoral thesis explores the use of Big Data in analysing fiscal information collected from electronic cash registers (ECRs), aiming to improve tax compliance, optimize revenue, and support informed policy formulation. The thesis begins with an Introduction presenting the context and motivation of the research, highlighting the significance of using Big Data to advance knowledge in the fiscal and technological fields. The objectives of the thesis are clearly defined, focusing on the development of an efficient platform for collecting and analysing fiscal data.

The chapter on the Current State of Research provides a general overview of the Big Data concept, its analysis techniques, and technologies, detailing its applications in the Romanian context of fiscal information. Case studies are reviewed, along with the specific characteristics of data collected from ECRs, emphasizing existing challenges and opportunities. The chapter also includes a detailed presentation of the fiscal data lifecycle, from collection, preprocessing, and cleaning, to analysis, visualization, and reporting.

The Methodology described in the third chapter outlines the steps of data collection, processing, and analysis, highlighting the tools and techniques used, as well as the methodological limitations encountered, including those related to the Romanian context. The evaluation and validation of the methodology are based on criteria of data validity and reliability, providing a robust framework for analysing fiscal information.

The chapter on Architecture and Design Specifications of the Data Collection and Analysis Platform explains in detail the types of information managed and the technologies used, including the modular architecture that facilitates data collection, transmission, preprocessing, and integration. The platform is designed to efficiently handle large data flows and ensure the validation and clear presentation of relevant information.

In the Platform Implementation chapter, the development and implementation of the platform's modules are discussed, including data reception, preprocessing, integration, analysis, and results publication. Each module is detailed in terms of functionality and contributions to the overall fiscal data analysis flow.

The Conclusions synthesize the thesis's main achievements, highlighting the original contributions, which include the development of an innovative architecture for the collection and analysis of fiscal data and demonstrating the impact of Big Data in improving tax compliance and optimizing revenues. Research gaps are also identified, and future directions are proposed for advanced exploration of Big Data technologies in the fiscal context. The list of original works reflects the author's commitment to disseminating research results and advancing knowledge in this emerging field.