

# **MICRO DATA MODEL ARCHITECTURE FOR AML SCORING RULE ENGINES**

Walakulu Arachchige Hasitha Maduranga

(199344L)

MSc in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

June 2022

# **MICRO DATA MODEL ARCHITECTURE FOR AML SCORING RULE ENGINES**

Walakulu Arachchige Hasitha Maduranga

(199344L)

Thesis submitted in partial fulfillment of the requirements for the degree of  
MSc in Computer Science Specializing in Software Architecture

Department of Computer Science and Engineering

University of Moratuwa  
Sri Lanka

June 2022

## **Declaration**

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature : ----- Date: -----

The above candidate has carried out research for the Masters thesis/ Dissertation under my supervision.

Name of the supervisor : Eng. Prof. Indika Perera

Signature of the supervisor : ----- Date: -----

## Abstract

Online and mobile banking have become a primary service of today's banking and financial sector. Clients could do their primary transactional jobs without physically appearing on the bank. This facility is 24x7 available. So, detection of money laundering activities based on transactional data analysis is a key challengeable area in today's banking and financial sector.

Businesses are trying to prevent money laundering activities by applying rule-based techniques to the real time operational transactions which could not completely cure the problem because higher constraints on the operational transaction could inconvenience the legal customer base and lose the customer satisfaction over the time. So, the near-real time and traditional data warehousing approaches with post detection techniques becomes the most common approach to detect money laundering activities in today's banking and financial context.

Traditional data warehousing approaches loaded data from operational or transactional systems on a weekly or nightly basis. Near real-time and real-time data warehouse approaches use real-time ETL tools to load data into the data warehouse in predefined shorter time intervals which preserve a gap with real-time transactional data. In addition to that, running anomaly detection engines (rule based or machine learning models) on top of those massive amounts of data (either OLTP databases or warehouse database) will take another considerable time due to higher velocity of data. So, identifying money launderers by analyzing post detection techniques causes higher risk to the financial system because the money launderer may leave the financial system before the money launderer catches.

This report introduce a novel **data modelling architecture** named "Micro Data Model Architecture" and an associated supporting tool named "Micro Temporal Database Generator" for "scoring rule engines" to detect financial fraudulent activities earlier by removing the burden on operational data sources.

## **Acknowledgement**

First, I must thank the Chancellor, Vice-Chancellor, and the rest of the management of the University of Moratuwa for giving me an opportunity to continue my higher studies in Computer Science with the guidance provided by the Department of Computer Science and Engineering University of Moratuwa. And also, thank them for continuously keeping the higher standard of the MSC program by keeping its curriculum up to date.

My deepest gratitude goes to my project supervisor, Prof.Indika Perera, for being interested in my research idea and constant support and guidelines rendered to me, right from the beginning, till the submission of the thesis report.

Further I thanks for the global industrial professionals, Mr.Boris Bialek, former Chief Architect of "Global Head of Banking Platform" in Switzerland (he is currently bare the post: Head of Innovation, EMEA at MongoDB) who shares his ideas and experience on real time transaction processing related to modern technology stack and allocate his valuable time to sharing his technical experiences with me. And, I would thank for management team of Creative Software Pvt.Ltd to giving me flexibility to studying the mastering program while employing. Also, I am grateful to all my teachers and my batch mates in MSC Program who supported and encouraged me throughout the completion of this project.

Finally, I express my deepest appreciation to my dearest wife and parents for their encouragement and support given throughout this complete.

## Table of Contents

1	CHAPTER 1 - INTRODUCTION TO MICRO DATA MODEL ARCHITECTURE FOR AML SCORING RULES ENGINES	1
1.1	Introduction	1
1.2	Background & Motivation	2
1.3	Problem in Brief	3
1.4	Aims and Objectives	4
1.5	Proposed Solution	5
1.6	Rest of the Report	7
1.7	Summary	7
2	CHAPTER 2 - LITERATURE REVIEW	9
2.1	Introduction	9
2.2	Banking Operations in the Modern Era	9
2.3	Propagation of the Money Laundering Problems	10
2.4	Anti-Money Laundering Solutions	12
2.4.1	AML Prevention Solutions/Techniques	13
2.4.2	AML Detection Solutions/Techniques	14
2.4.3	Hybrid Solutions/Techniques	15
2.5	Anti-Money Laundering Control Layered Framework	15
2.5.1	Real-time Controlling Layers	17
2.5.2	Near Real-time Controlling Layers	18
2.5.3	Offline Controlling Layers	18
2.6	Conventional and Temporal Databases	19
2.6.1	Conventional Databases	19
2.6.2	Temporal Databases	19
2.6.3	Relational Support for Temporal Databases	20
2.6.4	NoSQL Support for Temporal Database	20
2.7	Database Caching	21
2.7.1	What is caching?	21

2.7.2	Database Caching	21
2.8	Problem of Real World Transaction Data	24
2.8.1	Intrinsic Private Nature of Financial Transaction Data	24
2.8.2	Lack of Financial Transaction Data Quality	25
2.8.3	Financial Data Expose Can Cause Indirect Damages to Financial Institute	25
2.9	Synthetic Financial Transaction Data	25
2.9.1	PaySim Transaction Data	26
2.9.2	AMLSim Transaction Data	28
2.9.2.1	Supporting Typologies	29
2.9.2.2	Data Generate Process	33
2.10	Summary	35
3	CHAPTER 3 - METHODOLOGY	37
3.1	Introduction	37
3.2	Factors to Consider	37
3.2.1	Business Factors Force for Design Considerations	37
1.1.1	Technical Factors Force for Design Considerations	38
3.3	Micro Data Model Architecture In-Depth	40
3.4	Architectural Modeling	41
3.4.1	Scoring Rule Organization	42
3.4.2	Data Model Organization	44
3.5	Data Injection & Remove Process	48
3.6	Summary	49
4	CHAPTER 4 –MICRO TEMPORAL DATABASE GENERATOR	51
4.1	Introduction	51
4.2	Introduction to Micro Temporal Database Generator	51
4.3	Architecture	53
4.3.1	Rest API	54
4.3.2	Template Schema Mapper	54
4.3.3	Database Generator	56
4.4	Extendibility of Design	57
4.5	Model Validation & Testing	58
4.5.1	Generate Micro Temporal Databases	59
4.5.2	Data Design	60
4.5.3	Data Preprocessing & Loading	62

4.5.4	Scoring Rule Design	63
4.5.5	Model Validate	64
4.6	Summary	82
5	CHAPTER 5 – Discussion	83
5.1	Study Outcome	83
5.2	Experimental Limitations	84
5.3	Future Work	84
5.4	Summary	84
6	REFERENCES	85
7	APPENDIX A: ANTI-MONEY LAUNDERING LAWS AND REGULATIONS IN GLOBE	92
8	APPENDIX B: ANTI-MONEY LAUNDERING PATTERNS AND SAMPLE SCENARIOS (MOBILE BANKING)	95
9	APPENDIX C: GUIDELINES FOR MICRO TEMPORAL DATABASE GENERATOR	102
10	APPENDIX D: DATA PREPROCESSING AND LOADING FOR OLTP AND TEMPORAL DATABASES	115
11	APPENDIX E: MODEL VALIDATION TEST RESULTS SUMMARY	118



## List of Figures

Figure 1.1: Banking System Evolution Sketch over Time.....	1
Figure 1.2: Rough sketch Of Proposed Solution.....	6
Figure 2.1: AML Solution Approaches.....	13
Figure 2.2: Extended version of layered control flow of a fraud detection system.....	16
Figure 2.3:Bi-Temporal Data in JSON [Monger, Morgan & Mata-Toledo, Ramon & Gupta, Pranshu. (2012)].....	21
Figure 3.1: Rule Group Specialization.....	43
Figure 4.1: Generate Temporal Database Sample Request & Response .....	59
Figure 4.2: Generated Accounts.....	62
Figure 4.3: Generated Transactions .....	62
Figure 4.4: OLTP Database Schema Model.....	63
Figure 4.5: Temporal Database Schema Model.....	63
Figure 9.1: OLTP Schema Model .....	103
Figure 9.2: Schema Mapping File.....	106
Figure 9.3 : Temporal Schema Model.....	107
Figure 9.4: Create Temporal Database as 7 Day Data Cache .....	108
Figure 9.5: Create Temporal Database as 14 Day Data Cache .....	108
Figure 9.6: Create Temporal Database As 28 Day Data Cache .....	109
Figure 9.7: Verify Database .....	109
Figure 9.8: Verify Tables .....	110
Figure 9.9: Verify Events.....	110
Figure 9.10: Temporal Schema Script Location .....	110
Figure 9.11: Temporal Schema Generator Template Contract .....	114
Figure 10.1: Data Processing File Structure.....	115
Figure 10.2: SQL Data Insert Scripts for OLTP Database.....	115
Figure 10.3: SQL Data Insert Scripts for Temporal Database .....	116
Figure 10.4: Data Loading Configuration File.....	117
Figure 10.5: Data Loading Started.....	117

## List of Tables

Table 2.1:Money Laundering Scenario Comparison of Traditional Banking vs Modern Banking .....	12
Table 2.2:PaySim Supported Transaction Types .....	27
Table 4.1: Transaction Types Configuration Details .....	60
Table 4.2: Account Details Configuration .....	60
Table 4.3: Alert Pattern Configuration.....	61
Table 4.4: Scoring Rule 1 - Experiment 1 - Summary .....	68
Table 4.5: Scoring Rule 2 - Experiment 1 - Summary .....	73
Table 4.6: Scoring Rule 3 - Experiment 1 - Summary .....	76

Table 4.7: Scoring Rule 4 - Experiment 1 - Summary.....	81
Table 9.1: Schema Mapping Guidelines .....	102
Table 9.2 : Schema Mapping Attribute Definitions .....	106
Table 11.1: Scoring Rule 1 - Experiment 1 - Summary.....	118
Table 11.2: Scoring Rule 2 - Experiment 1 - Summary.....	118
Table 11.3: Scoring Rule 3 - Experiment 1 - Summary.....	119
Table 11.4: Scoring Rule 4 - Experiment 1 - Summary.....	120
Table 11.5: Scoring Rule 1 - Experiment 2 - Summary.....	121
Table 11.6: Scoring Rule 2 - Experiment 2 - Summary.....	121
Table 11.7: Scoring Rule 3 - Experiment 2 - Summary.....	122
Table 11.8: Scoring Rule 4 - Experiment 2 - Summary.....	122
Table 11.9: Scoring Rule 1 - Experiment 3 - Summary.....	123
Table 11.10: Scoring Rule 2 - Experiment 3 - Summary.....	124
Table 11.11: Scoring Rule 3 - Experiment 3 - Summary.....	125
Table 11.12: Scoring Rule 4 - Experiment 3 - Summary.....	125
Table 11.13: Scoring Rule 1 - Experiment 4 - Summary.....	126
Table 11.14: Scoring Rule 2 - Experiment 4 - Summary.....	127
Table 11.15: Scoring Rule 3 - Experiment 4 - Summary.....	127
Table 11.16: Scoring Rule 4 - Experiment 4 - Summary.....	128
Table 11.17: Scoring Rule 1 - Experiment 5 - Summary.....	129
Table 11.18: Scoring Rule 2 - Experiment 5 - Summary.....	129
Table 11.19: Scoring Rule 3 - Experiment 5 - Summary.....	130
Table 11.20: Scoring Rule 4 - Experiment 5 - Summary.....	130

## List of Abbreviations

AML	Anti-Money Laundering
FATF	Financial Action Task Force
FCA	Financial Conduct Authority (UK)
5AMLD	Fifth Anti-Money Laundering Directive (European Union)
6AMLD	Sixth Anti-Money Laundering Directive (European Union)
ICA	International Compliance Association
OLTP	Online Transaction Processing
OLAP	Online Analytical Processing
IMoLIN	International Money Laundering Layering Network
ICA	International Compliance Association
CDC	Change Data Capture
ACID	Atomicity,Consistency,Isolation,Durability
SC	Scoring Rule