

INSIGHT STUDY OF VARIOUS MODEL USED FOR ONLINE ANALYTICAL PROCESS

A Thesis

Submitted towards the Requirement for the Award of Degree of

DOCTOR OF PHILOSOPHY

IN

COMPUTER SCIENCE AND APPLICATION

Under the Faculty of Computer Science and Application

By

ANJANA YADAV

(Enrollment No – 161595506208)

Under the Supervision of

Dr. Balveer Singh

Professor, Department of Computer Science & Application

P.K. UNIVERSITY



2023

P.K. University

NH-27, Village. Thanra (P.O. - DINARA),

Shivpurii MP. India—473665

www.pkuniversity.edu.in

CERTIFICATE OF THE SUPERVISOR

This is to certify that the work entitled "**Inside Study of Various Model used for Online Analytical Process**" is a piece of research Work done by Smt. Anjana Yadav Under my Guidance and Supervision for the degree of Doctor of Philosophy of Computer Science and Application, P.K. University (M.P) India.

I certify that the candidate has put an attendance of more than 240 day with me. To the best of my Knowledge and belief the thesis:

I – Embodies the work of the candidate himself/herself.

II – Has duly been completed.

III – Fulfill the requirement of the ordinance relating to the Ph.D. degree of the University.

Signature of the Supervisor

Date:



P.K. University

(University established under section 2f of UGC act 1956 vide mp government
act no 17 of 2015)

NH-27, Village. Thanra (P.O.-DINARA), Shivpurii M.P. India—473665

DECLARATION BY THE CANDIDATE

I declare that the thesis entitled "**Inside Study of Various Model used for Online Analytical Process**" is my own work conducted under the supervision of **Dr. Balveer Singh** Supervisor at P. K. University, Shivpuri Approved by Research Degree Committee. I have put more than 240 days of attendance with Supervisor at the center.

I further declare that to the best of my knowledge the thesis does not contain my part of any work has been submitted for the award of any degree either in this University or in any other University Without proper citation.

Signature of the candidate

Date:

Place:

FORWARDING LETTER OF HEAD OF INSTITUTION

The Ph.D. thesis entitled "**Inside Study of Various Model used for Online Analytical Process**" Submitted by **Smt. Anjana Yadav** is forwarded to the university in six copies. The candidate has paid the necessary fees and there are no dues outstanding against him/her.

Name
.....

Seal

Date:

Place:

(Signature of Head of Institution where
the Candidate was registered for Ph.D degree)

Signature of the Supervisor Date:

Date:

Address:

Place:

.....

.....

ACKNOWLEDGEMENT

I take this opportunity to express my gratitude and sincere thanks to **Dr. Balveer Singh**, Professor Department of Computer Science and Application ,P. K. University, Shivpuri, for his able guidance, valuable help enthusiastic attitude and suggestions throughout the period of my research work. I am very fortunate for having the opportunity to work under him. He had been a perennial source of inspiration and his guidance helped me to find the ways through hurdles.

I am indeed grateful to **Dr. Jitendra Kumar Mishra**, Director, P. K. University, Shivpuri, for providing the necessary facilities, infrastructure and encouragement for completion of thesis work.

I would not complete my duty if I omit to pay my sincere gratitude to all the persons, friends and relatives, who have directly or indirectly helped me in the completion of this work.

ANJANA YADAV
Enrollment No – 161595506208

“LIST OF ABBREVIATION”

ACO	-	Ant Colony Optimization
ALO	-	Ant Lion Optimization
CF	-	Configuration
CS	-	Consistency
DFOC	-	Dragon Fly Optimization based Clustering
ER	-	Entity Relationship
ETL	-	Extraction Transformation Loading
FA	-	Foodstuff sources Appeal
FFO	-	Fruit Fly Optimization
GA	-	Genetic Algorithm
GWO	-	Grey Wolf Optimization
HOLAP	-	Hybrid Online Analytical Process
HTML	-	Hyper Text Mark-up Language
IoMT	-	Internet of Medical Things
IoT	-	Internet of Thing
IT	-	Information Technology
K-SSOC	-	KMeans-Salp Swarm Optimization based Clustering
KPIs	-	Key Performance Indicators
MAC	-	Multidimensional Aggregation Cube
MATLAB	-	Matrix Laboratory
MOLAP	-	Multi-dimensional Online Analytical Process
MS	-	Maintenance Spending
OE	-	Opponent Escaping
OLAP	-	Online Analytical Process
OLTP	-	Online Transactional Processing
PDF	-	Portable Document Format
PSO	-	Particle Swarm Optimization
RDBMS	-	Relational Data Base Management System
ROLAP	-	Relational Online Analytical Process
SBI	-	Social Business Intelligence

SQL	-	Structured Query Language
SR	-	Standard Severance
UML	-	Unified Modeling Language
WMS	-	Whole Maintenance Spending
WQS	-	Whole Query Spending
WSF	-	Whole Maintenance Function
XML	-	Extensible Mark-up Language

“LIST OF TABLES”

Table No.	Name of Table	Page No.
2.1	Comparative Analysis of OLAP based Research Works	26
6.1	Multidimensional Clinical Datasets	64
6.2	Results for Cancer Dataset	67
6.3	Results for Cryotherapy Dataset	68
6.4	Results for Liver Patients dataset	68
6.5	Results for Heart Patients Dataset	68
6.6	Results for average rank for all Datasets based on Intra-Cluster Distance	69
7.1	Multidimensional Datasets	77
7.2	Outcomes for Datasets	81
7.3	Average Ranking of all approaches for total Datasets in terms of total of intra-cluster distances mean values	82
8.1	Multidimensional Datasets	88
8.2	Average Ranking (Intra-cluster Distance) for Multidimensional Datasets	89
8.3	F-Measure for Multidimensional Datasets	91
8.4	Purity Index for Multidimensional Datasets	92
8.5	Standard Deviation for Multidimensional Datasets	94

“LIST OF FIGURES”

Figures No.	Figure Name	Page No.
1.1	MOLAP Architecture	3
1.2	ROLAP Architecture	4
1.3	HOLAP Architecture	4
1.4	OLAP Analysis	7
1.5	Flow Chart of OLAP Multidimensional Model Cube Selection	17
3.1	Lattice Structure	37
3.2(a)	Implementation in MATLAB	39
3.2(b)	Implementation in MATLAB	39
3.2(c)	Implementation in MATLAB	40
3.3	Query Processing Expenditure for PSO and FFO approaches (3 Dimensions)	40
3.4	Query Processing Expenditure for PSO and FFO approaches (4 Dimensions)	41
3.5	Query Processing Expenditure for PSO and FFO approaches (Identical Frequencies)	41
3.6	Query Processing Expenditure for PSO and FFO approaches (Arbitrary Frequencies)	42
4.1	Lattice Structure	47
4.2(a)	Implementation in MATLAB	49
4.2(b)	Implementation in MATLAB	49
4.2(c)	Implementation in MATLAB	50
4.3	Query Dispensation Expenses on 3 Dimensions	50
4.4	Query Dispensation Expenses on 4 Dimensions	51
4.5	Query Dispensation Expenses on 5 Dimensions	52
4.6	Query Dispensation Expenses for Uniform Frequencies	53
4.7	Query Dispensation Expenses for Capricious Frequencies	53
5.1	Query Dispensation Expenses on 3 Dimensions	56
5.2	Query Dispensation Expenses on 4 Dimensions	56
5.3	Query Dispensation Expenses on 5 Dimensions	57
5.4	Query Dispensation Expenses for Uniform Frequencies	58

5.5	Query Dispensation Expenses for Capricious Frequencies	58
6.1(a)	Implementation in MATLAB	64
6.1(b)	Implementation in MATLAB	65
6.1(c)	Implementation in MATLAB	65
6.1(d)	Implementation in MATLAB	66
6.2	Average Rank for all datasets based on Intra-cluster Distance	69
6.3	F-Measure for all datasets	70
6.4	Purity Index for all datasets	70
6.5	Standard Deviation for all datasets	71
7.1(a)	Implementation in MATLAB	78
7.1(b)	Implementation in MATLAB	78
7.1(c)	Implementation in MATLAB	79
7.1(d)	Implementation in MATLAB	79
7.2	Average Rank for six datasets in terms of Intra-cluster Distance	83
7.3	Purity Index for total six datasets	83
7.4	F-Measure for total six datasets	84
7.5	Standard Deviation for total six datasets	85
8.1	Average Rank for ten datasets in terms of Intra-cluster Distance	90
8.2	F-Measure for total ten Multidimensional datasets	92
8.3	Purity Index for total Multidimensional datasets	93
8.4	Standard Deviation for total ten Multidimensional datasets	95

COURSE WORK CERTIFICATE

This is to certify that **AnjanaYadav**(Enrollment No.161595506208) D/O Mr.J.C.Yadav,Scholar of the Ph.D in Department Of Computer Science and Application Under the faculty of Computer Science and Application has successfully passed the course work examination with 'A' grade from P.K.University, Shivpuri (M.P.).

Registrar
P.K.University
Shivpuri (M.P.)

PLAGIARISM REPORT FROM THE CENTRAL LIBRARY

CERTIFICATE OF PLAGIARISM REPORT

1. Name of the Research Scholar : Anjana Yadav
2. Course of Study : Doctor of Philosophy (Ph.D)
3. Title of the Thesis : Inside Study of Various Model used for Online
Analytical Process
4. Name of the Supervisor : Dr. Balveer Singh
5. Department : Computer Science and Application
6. Subject : Computer Science
7. Acceptable Maximum Limit : 10% (As per UGC Norms)
8. Percentage of Similarity of
Contents Identified : 0%
9. Software Used : Original (Formerly URKUND)
10. Date of Verification :

**Signature of Our original Coordinator
(Librarian, Central Library)
P.K.University, Shimpuri (M.P.)**

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Inside Study of Various Model used for Online Analytical Process

Candidate's Name: Anjana Yadav

COPYRIGHT TRANSFER TO

The undersigned hereby assigns to the P.K. University, Shivpuri all copyrights that exist in and for the above thesis submitted for the award of the Ph.D. degree.

Date:

Anjana Yadav

Name of the scholar

Note: However, the author may reproduce/publish or authorize others to reproduce, material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the University's copyright notice are indicated.

"ABSTRACT"

The Online Analytical Processing (OLAP) based Multidimensional examination hassles for several stockpiling magnificence over huge data. For as much to recognize queries answering time companionable by OLAP framework users and understanding entire business perceive mandatory, OLAP data is structured as a data cube (a multidimensional model). The OLAP queries are responded in speedy and steady time by utilizing the cube materialization for assessments takers. But, this also involves unendurable expenses, regarding to stockpile memory and period, and as a data depot, OLAP has an average dimension and dimensionality which is to be significant on query processing. Consequently, cube assortment has got to be finished motivating to diminish inquiry management expenses, maintaining as a restraint the materializing gap. Several techniques and heuristics like deviationist and insatiable algorithms have been utilized to offer an estimated result. In this work, a Fruit Fly Optimization (FFO) approach is implemented in a lattice structure to obtain an optimal materialized data cube for reducing the query processing expenses. The results illustrate that FFO generates better performance than Particle Swarm Optimization (PSO) in terms of frequency and number of dimensions.

The data cube assessments dependent on Online Analytical Processing (OLAP) trouble for numerous depositing splendors over broad information. In favor of appreciating question answering era pleasant with OLAP skeleton patrons and allowing complete industry organized notice compulsory, OLAP information is organized as a data cube model. The OLAP questions are answered in rapid and sturdy time by exploiting the cube embodiment for appraisals buyers. Until now this moreover insets insupportable charge, concerning to accumulation remembrance and time, yet as a data storage area had a typical length and extent which will be influential on stimulating procedure. Thus, cube classification has visited to be refined fascinating to moderate question managing charge, preserving as a control the materializing breach. Numerous strategies and heuristics like divergence and voracious approaches have been exploited to suggest a vague solution. Here, a Grey Wolf Optimization (GWO) strategy is exploited in a lattice structure for finding the best data cube to decrease the question processing charge. The outputs describe the superior efficiency of GWO against GA, PSO and ALO based on total dimensions and frequency.

Medicine is a fresh way to utilize for curing, analyzing and detecting the diseases through data clustering with OLAP (Online Analytical Processing). The large amount of multidimensional clinical data is reduced the efficiency of OLAP query processing by enhancing the query accessing time. Hence, the performance of OLAP model is improved by using data clustering in which huge data is divided into several groups (clusters) with cluster heads to achieve fast query processing in least time. In this chapter, a Dragon Fly Optimization based Clustering (DFOC) approach is proposed to enhance the efficiency of data clustering by generating optimal clusters from multidimensional clinical data for OLAP. The results are evaluated on MATLAB 2019a tool and shown the better performance of DFOC against other clustering methods ACO, GA and K-Means in terms of intra-cluster distance, purity index, F-measure, and standard deviation.

The performance of query processing over OLAP (Online Analytical Processing) model is decreased due to higher query access time for huge multidimensional data. Therefore, the clustering is introduced to improve the OLAP model efficiency by getting quick query processing because of dividing the large data into various clusters. The K-Means is a famous technique of clustering the data into groups to solve various real life issues. However, K-Means has some drawbacks like sensitivity to primary centroid assortment in cluster and local optimum convergence. Hence, a KMeans-Salp Swarm Optimization based Clustering (K-SSOC) is implemented to improve the performance of K-Means by providing optimal clustering over huge OLAP multidimensional data. The outcomes are obtained on MATLAB 2019a environment based on the parameter purity index, standard deviation, F-measure, intra-cluster distance and running time complexity over 1000 iterations. The results illustrate the superior performance of K-SSOC against K-Means, ACO and PSO over total six multidimensional datasets based on parameters.

“TABLE OF CONTENTS”

Chapter	Title	Page No.
Chapter 1	Introduction	1-21
1.1	Introduction	1
1.2	OLAP Models	2
1.2.1	MOLAP Model	2
1.2.2	ROLAP Model	3
1.2.3	HOLAP Model	4
1.3	Data Warehouse	5
1.4	Online Transactional Processing	7
1.5	Optimization Methods & its Components	8
1.6	Optimization Techniques	9
1.7	Open Research Issues	10
1.8	Key Performance Factors	11
1.8.1	Intra-cluster distance	11
1.8.2	Purity Index	11
1.8.3	F-Measure	11
1.8.4	Standard deviation	12
1.9	Motivation	12
1.10	Problem Identification	13
1.11	Contribution of Research	14
1.12	Research Objectives	15
1.13	Scope of Work	16
1.14	The Description of OLAP Multidimensional Model for Cube Selection	18
1.15	The Outcomes of This Research Work	18
1.16	Organization of Thesis	19
1.17	Summary and Discussion	20
Chapter 2	State of the Art: Review	22-33
2.1	Introduction	22
2.2	Literature Review	22
2.3	Summary and Discussion	32
Chapter 3	Selection of OLAP Materialized Cube by using a Fruit Fly Optimization (FFO) Approach: a Multidimensional Data Model	34-43
3.1	Introduction	34

	3.2 The Fruit Fly Optimization (FFO) approach for selection of OLAP Materialized Cube (a Multidimensional Data Model)	35
	3.2.1 FFO Approach	35
	3.2.2 Lattice Structure	36
	3.2.3 Cube Selection using FFO approach	37
	3.3 Result and Analysis	38
	3.4 Summary and Discussion	42
	3.5 Limitation of FFO Approach	43
Chapter 4	A Grey Wolf Optimization (GWO) based Cube Selection in OLAP Data Model	44-54
	4.1 Introduction	44
	4.2 The Proposed Grey Wolf Optimization (GWO) based Cube Selection in OLAP Data Model	45
	4.2.1 GWO Approach	45
	4.2.2 Lattice Structure	46
	4.2.3 GWO approach for Cube Selection	47
	4.3 Result Analysis	48
	4.4 Summary and Discussion	54
Chapter 5	Comparative Analysis of FFO and GWO Approaches	55-60
	5.1 Introduction	55
	5.2 Comparative Analysis of FFO and GWO Approaches	55
	5.3 Summary and Discussion	59
Chapter 6	Improving the Performance of Multidimensional Clinica; Data for OLAP using an Optimized Data Clustering Approach	61-72
	6.1 Introduction	61
	6.2 Dragon Fly Optimization based Clustering (DFOC) Approach	62
	6.2.1 DFOC Approach	62
	6.2.2 Multidimensional Clinical Datasets	64
	6.3 Result and Analysis	64
	6.3.1 Intra-cluster distance	66
	6.3.2 Purity Index	66
	6.3.3 F-Measure	67
	6.3.4 Standard Deviation	67
	6.4 Summary and Discussion	71
	6.5 Limitation of DFOC approach	72
Chapter 7	Improve the Performance of Multidimensional Data for OLAP by using an Optimization Approach	73-87
	7.1 Introduction	73
	7.2 KMeans-Salp Swarm Optimization based Clustering (K-SSOC) Approach	74

7.2.1	K-SSOC Approach	74
7.2.2	Multidimensional Datasets	77
7.3	Result and Discussion	78
7.3.1	Intra-cluster Distance	79
7.3.2	Purity Index	80
7.3.3	F-Measure	80
7.3.4	Standard Deviation	81
7.3.5	Time Complexity	85
7.4	Summary and Discussion	86
Chapter 8	Comparative Analysis of DFOC and K-SSOC Approaches	88-96
8.1	Introduction	88
8.2	Multidimensional Datasets	88
8.3	Comparative Analysis of DFOC and K_SSOC Approaches	89
8.4	Time Complexity	96
8.5	Summary and Discussion	96
Chapter 9	Conclusion and Future Direction	97-98
9.1	Conclusions	97
9.2	Future Direction	98
References		99
Appendix I:	List of Papers Published/ Presented/ Communicated By The Candidate	109
Appendix II	Plagiarism Report	110