Contents

1 Introduction **1**

1.1 Data Warehousing and Materialized Views 2

1.1.1 Multidimensional data model 3

1.1.2 Aggregations of data as data warehouse views 5

1.1.3 Materialized views for efficient computation of data cubes . 7

1.1.4 Materializing views in RDBMS technology 7

1.1.5 Common SQL sub-expressions of SQL queries as materialized views 8

1.1.6 Materializing results of common sub-expressions in Big data framework 11

1.2 Multi-Objective Optimization to Select Views for Materializing . . . 12

1.2.1 The view selection problem for materializing 12

1.2.2 The costs to be minimized and trade-offs 13

1.2.3 Multi-Objective Optimization (MOO) 13

1.2.4 Selecting views by multi-objective optimization technique . . 14

1.3 Motivation of this Research 15

1.4 Research Objectives 15

1.5 Contributions 17

1.6 Organization of the Thesis 17

2 Materialized View Selection in Data Warehouses: Approaches, Issues and Challenges **19**

2.1 Introduction 19

2.2 Representations of views in Data Warehouses 20

2.2.1 Multidimensional lattice representation of views 20

2.2.2 AND-OR view graph representation of queries and views . . 22

2.2.3 Optimal multiple query execution plan based graphical representation 24

2.2.4 Query - Attribute matrix representation 26

2.2.5 Associated issues in different representations 26

2.3 Existing View Selection Techniques 28

2.3.1 Greedy algorithmic approaches 28

2.3.2 Stochastic algorithmic approaches 33

2.3.3 Data mining based approaches 39

2.4 A Brief Discussion on Existing Approaches 42

2.4.1 Issues and challenges 44

2.5 Discussion 45

3 Multi-Objective Differential Evolution Algorithm for Selecting Views to Materialize 47

3.1 Introduction 47

3.1.1 Motivation 47

3.1.2 Contribution 48

3.2 The View Selection for Materializing as a Multi-objective Optimization Problem 49

3.2.1 DAG representation of multiple query processing plan 50

3.2.2 The cost model 51

3.2.3 Multi-objective optimization 53

3.2.4 The view selection problem as multi-objective optimization problem representation 54

3.2.5 Solution representation 55

3.3 Multi-objective Differential Evolution Algorithm for Selecting Views to Materialize in Data Warehouse 56

3.3.1 The Differential Evolution (DE) algorithm 56

3.3.2 Differential Evolution algorithm adapted with binary encoded data 57

3.3.3 Multi-objective DE 60

3.3.4 Multi-objective DE with binary encoded data for view selection : MODE-BE 62

3.3.5 Complexity analysis 64

3.3.6 Convergence 64

3.4 Experimentation and Observations 67

3.4.1 The test-bed used 67

3.4.2 Control parameters 67

3.4.3 Observations 68

3.5 Discussion 70

4 Materialized View Selection by Evolutionary Algorithm for Big Data Query Processing 75

4.1 Introduction 75

4.1.1 Materialized views and materialized queries in Big data . . . 76

4.1.2 View selection for materializing in Big data 77

4.1.3 Contribution 77

4.2 The Problem of Selecting Views for Materializing in Big data Framework 78

4.2.1 The cost model and problem definition 78

4.3 View Selection in Big data Systems as Multi-Objective Optimization Problem 81

4.3.1 Simple problem representation 81

4.3.2 Scalability 82

4.3.3 Well defined objectives 82

4.4 Multi-Objective Evolutionary Algorithm for View Selection to Materialize in Big data . . . 83

4.4.1 Multi-objective DE with binary encoded solutions for Big data view selection . . . 83

4.4.2 Implementing NSGA-II for view selection . . . 87

4.5 Experimentation and Observations . . . 88

4.5.1 Experimental setup . . . 89

4.5.2 Parameters used . . . 92

4.5.3 Results and observations . . . 98

4.6 Discussion . . . 99

5 Multi-Objective Simulated Annealing Algorithm in Big data View Selection for Materializing 101

5.1 Introduction . . . 101

5.1.1 Motivation . . . 101

5.1.2 SA for multi-objective optimization . . . 102

5.1.3 Contribution . . . 103

5.2 Dominance based Energy Functions in Multi-Objective Simulated Annealing Algorithm . . . 103

5.2.1 Energy function in terms of number of dominating solutions 104

5.2.2 Energy function in terms of amount of domination . . . 106

5.3 Representation of the View Selection Problem for Applying AMOSA 110

5.3.1 Representing view selection problem for Big data . . . 110

5.3.2 Solution representation . . . 111

5.4 AMOSA for Materialized View Selection . . . 112

5.4.1 Initializing the archive of solutions . . . 112

5.4.2 Enforcing diversity in solution space for restricting the archive size . . . 112

5.4.3 The main process of AMOSA-MVS . . . 114

5.4.4 Parameter selection 116
5.4.5 Complexity analysis 119
5.4.6 Convergence 121
5.5 Experimental Results and Performance Analysis 122
5.5.1 Experimental setup and test data sets 122
5.5.2 Experimentation and results 125
5.5.3 Comparison measures 126
5.5.4 Comparative analysis 132
5.6 Discussion 134

6 Conclusion and Future Direction **141**

6.1 Conclusions 141
6.2 Future Directions 143

Bibliography **145**

Publications based on the Thesis Works **153**