UNIVERSITI PUTRA MALAYSIA

DOMINANCE RELATIONSHIP-BASED SKYLINE QUERY FRAMEWORK OVER DYNAMIC AND INCOMPLETE DATABASE

GHAZALEH BABANEJAD DEHAKI

FSKTM 2021 9

DOMINANCE RELATIONSHIP-BASED SKYLINE QUERY FRAMEWORK OVER DYNAMIC AND INCOMPLETE DATABASE

By

GHAZALEH BABANEJAD DEHAKI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

October 2020

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

DOMINANCE RELATIONSHIP-BASED SKYLINE QUERY FRAMEWORK OVER DYNAMIC AND INCOMPLETE DATABASE

By

GHAZALEH BABANEJAD DEHAKI

October 2020

Chairman : Professor Hamidah Ibrahim, PhD
Faculty : Computer Science and Information Technology

Skyline queries rely on the notion of Pareto dominance, filter the data items by keeping only those data items that are the best, most preferred, also known as skylines, from a database to meet the user's preferences. Skyline query has been studied extensively and a significant number of skyline algorithms have been proposed, mostly attempt to resolve the optimisation problem that is mainly associated with a reduction in the processing time of skyline computations. In today's era, the presence of incomplete data in a database is inevitable. The skyline algorithms in such situation will have to deal with several issues besides the optimisation problem. The missing values in databases give a negative influence on the number of pairwise comparisons that needs to be performed between the data items. Moreover, the transitivity property of skylines is no longer hold. Cyclic dominance is another issue that needs to be tackled as it yields empty skyline results. Furthermore, databases are dynamic in nature in which their states change throughout the time. These changes are necessary as databases must reflect the current and latest information of the applications. The changes are normally achieved through data manipulation operations and data definition operations. The skylines derived before changes are made towards the initial database are no longer valid in the new state of the database. Utilising the existing skyline algorithms would require performing the algorithms on the new state of the database. However, computing the skylines over the entire database after changes are made is inefficient as not all the data items are affected by the changes.

In tackling the above stated issues, we propose a solution, named *DyIn-Skyline*, which consists of three main phases, namely: *Phase* I – processing skyline queries over the initial incomplete database, *Phase* II – processing skyline queries over a dynamic and incomplete database, in which the changing state of the database is due to a data manipulation operation(s) (insert, delete or update a data item(s)), and *Phase* III – processing skyline queries over a dynamic and incomplete database, in which the

changing state of the database is due to a data definition operation(s) (add or remove a dimension(s)). For each phase, a framework is proposed. The proposed framework in the *Phase* I consists of three main components, namely: *Data Grouping Builder* (*DGB*), *Bucket Skyline Identifier* (*BSI*), and *Final Skyline Identifier* (*FSI*). We have also introduced and designed three lists, namely: *Bucket Dominating* (*BDG*), *Bucket Dominated* (*BDD*), and *Domination History* (*DH*) to keep track of the dominating data items, dominated data items, and dominance relationships, respectively; this information is useful and is utilised by the *Phase* II and *Phase* III of the *DyIn-Skyline* solution. The framework of *Phase* II consists of three components, namely: *Skyline-Insert Identifier (S-II*), which derives a set of skylines after a data item(s) is inserted into a database, *Skyline-Delete Identifier* (*S-DI*), which derives a set of skylines after an existing data item(s) is deleted from a database, and *Skyline-Update Identifier* (*S-UI*), which produces a set of skylines after an existing data item(s) of a database is updated. Meanwhile, the framework of *Phase* III consists of two components, namely: *Skyline-Add Dimension Analyser* (*S-ADA*) which derives a set of skylines after a new dimension(s) is added to a database and *Skyline-Remove Dimension Analyser* (*S-RDA*) which derives a set of skylines after an existing dimension(s) is removed from a database.

Extensive experiments have been conducted to evaluate the performance and prove the efficiency of our proposed solution, *DyIn-Skyline*, in processing skyline queries over a dynamic and incomplete database. The performance results of *DyIn-Skyline* are compared to other existing works that are the closest to this research, namely: *ISkyline*, *SIDS*, and *Incoskyline*. In most cases, *DyIn-Skyline* shows a steady performance and achieves better performance with regard to the number of pairwise comparisons and processing time compared to the previous works. Unlike *ISkyline*, *SIDS*, and *Incoskyline* which derive skylines over the entire database after changes are made towards the database, i.e. the new state of the database, *DyIn-Skyline* avoids unnecessary skyline computations. It relies on the information saved in the following lists: *Bucket Dominating* (*BDG*), *Bucket Dominated* (*BDD*), and *Domination History* (*DH*) and focuses only on those data items that are affected by the changes.

DEDICATION

To my beloved father who always supports me and taught me how to be strong and hard worker

To my beloved late mother who taught me always have confidence on myself and live happy

This is for you mom and dad!

Ghazaleh

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENGIRAAN EFISIEN BAGI PERTANYAAN LATAR LANGIT KE ATAS PANGKALA DATA DINAMIK DAN TIDAK LENGKAP

Oleh

GHAZALEH BABANEJAD DEHAKI

Oktober 2020

Pengerusi : Profesor Hamidah Ibrahim, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Pertanyaan latar langit bergantung kepada tanggapan penguasaan Pareto, menyaring item data dengan hanya menyimpan item data yang terbaik, paling digemari, juga dikenali sebagai latar langit, daripada sebuah pangkalan data untuk memenuhi keutamaan pengguna. Pertanyaan latar langit telah dikaji secara meluas dan sebilangan besar algoritma latar langit telah dicadangkan, kebanyakannya berusaha untuk menyelesaikan masalah pengoptimuman yang berkait terutamanya dengan pengurangan di dalam masa pemprosesan pengiraan latar langit. Dalam era hari ini, kehadiran data tidak lengkap di dalam pangkalan data tidak dapat dielakkan. Algoritma latar langit di dalam situasi sedemikian akan perlu berurusan dengan beberapa isu di samping masalah pengoptimuman. Nilai yang hilang di dalam pangkalan data memberi pengaruh negatif ke atas bilangan perbandingan berpasangan yang perlu dilakukan antara item data. Lebih-lebih lagi, sifat transitiviti latar langit tidak lagi dipegang. Keutamaan kitaran adalah isu lain yang perlu diatasi kerana ia menghasilkan keputusan latar langit yang kosong. Lagi pun, pangkalan data adalah dinamik secara semula jadi di mana keadaannya berubah sepanjang masa. Perubahan ini adalah perlu kerana pangkalan data mesti mencerminkan maklumat semasa dan terkini aplikasi. Perubahan ini biasanya dicapai melalui operasi manipulasi data dan operasi definisi data. Latar langit yang diterbitkan sebelum perubahan dibuat ke atas pangkalan data asal tidak lagi sah di dalam keadaan baharu pangkalan data. Menggunakan algoritma latar langit sedia ada memerlukan pelaksanaan algoritma tersebut ke atas keadaan baharu pangkalan data. Namun begitu, mengira latar langit ke atas keseluruhan pangkalan data selepas perubahan dibuat adalah tidak efisien kerana tidak semua item data adalah terjejas dengan perubahan tersebut.

Di dalam mengatasi isu yang disebut di atas, kami mencadangkan satu penyelesaian, dinamakan *DyIn-Skyline*, yang mengandungi tiga fasa utama, iaitu: *Fasa* I – memproses pertanyaan latar langit ke atas pangkalan data tidak lengkap asal, *Fasa* II

– memproses pertanyaan latar langit ke atas pangkalan data dinamik dan tidak lengkap, di mana perubahan keadaan pangkalan data disebabkan operasi manipulasi data (masuk, padam, atau kemas kini item data), dan *Fasa* III – memproses pertanyaan latar langit ke atas pangkalan data dinamik dan tidak lengkap, di mana perubahan keadaan pangkalan data disebabkan operasi definisi data (tambah atau buang dimensi). Untuk setiap fasa, satu kerangka dicadangkan. Kerangka yang dicadangkan di dalam *Fasa* I mengandungi tiga komponen, iaitu: *Data Grouping Builder* (*DGB*), *Bucket Skyline Identifier* (*BSI*), dan *Final Skyline Identifier* (*FSI*). Kami juga telah memperkenalkan dan mereka bentuk tiga senarai, iaitu: *Bucket Dominating* (*BDG*), *Bucket Dominated* (*BDD*), dan *Domination History* (*DH*) untuk mengesan item data menguasai, item data dikuasi, dan pertalian penguasaan, masing-masing; maklumat ini adalah berguna dan digunakan oleh *Fasa* II and *Fasa* III penyelesaian *DyIn-Skyline*. Kerangka *Fasa* II mengandungi tiga komponen, iaitu: *Skyline-Insert Identifier* (*S-II*), yang menerbitkan satu set latar langit selepas item data dimasukkan ke dalam pangkalan data, *Skyline-Delete Identifier* (*S-DI*), yang menerbitkan satu set latar langit selepas item data sedia ada dipadamkan dari pangkalan data, dan *Skyline-Update Identifier* (*S-UI*), yang menghasilkan satu set latar langit selepas item data sedia ada pangkalan data dikemas kini. Sementara itu, kerangka *Fasa* III mengandungi dua komponen, iaitu: *Skyline-Add Dimension Analyser* (*S-ADA*) yang menerbitkan satu set latar langit selepas satu dimensi baharu ditambah ke dalam pangkalan data dan *Skyline-Remove Dimension Analyser* (*S-RDA*) yang menerbitkan satu set latar langit selepas dimensi sedia ada dibuang dari pangkalan data.

Experimen yang luas telah dijalankan untuk menilai prestasi dan membuktikan kecekapan penyelesaian cadangan kami, *DyIn-Skyline*, di dalam memproses pertanyaan latar langit ke atas pangkalan data dinamik dan tidak lengkap. Keputusan prestasi *DyIn-Skyline* dibandingkan dengan kerja sedia ada lain yang paling hampir dengan penyelidikan ini, iaitu: *ISkyline*, *SIDS*, dan *Incoskyline*. Dalam kebanyakan kes, *DyIn-Skyline* menunjukkan prestasi yang stabil dan mencapai prestasi yang lebih baik dengan mengambil kira bilangan perbandingan berpasangan dan masa pemprosesan berbanding dengan kerja sebelum ini. Tidak seperti *ISkyline*, *SIDS*, dan *Incoskyline* yang menerbitkan latar langit ke atas keseluruhan pangkalan data selepas perubahan dibuat ke atas pangkalan data, iaitu keadaan baharu pangkalan data, *DyIn-skyline* mengelak pengiraan latar langit yang tidak perlu. Ia bergantung kepada maklumat yang disimpan dalam senarai berikut: *Bucket Dominating* (*BDG*), *Bucket Dominated* (*BDD*), dan *Domination History* (*DH*) dan memfokus hanya pada item data yang terjejas oleh perubahan tersebut.

ACKNOWLEDGEMENTS

In the name of god, the most merciful and most compassionate. All praises to Allah and His blessing for the completion of this thesis. I thank God for all the opportunities, trials and strength that have been showered on me to finish my thesis. I experienced so much during this process, not only from the academic aspect but also from the aspect of personality. During the completion of my work I have received encouragement from several quarters and it is my pleasant duty to express my gratitude to all concerned.

In my journey towards this degree, I have found a teacher, a friend, an inspiration, a role model and a pillar of support in my Guide, Professor Dr. Hamidah Ibrahim. I would like to express my sincere gratitude to her for the continuous support of my Ph.D study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Assoc. Prof. Dr. Nor Izura Udzir, Assoc. Prof. Dr. Fatimah Sidi, and Dr. Ali Amer Alwan for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I also I would like to thank many people I have met during my stay in Malaysia for their help, enjoyable discussions and some good times. I delighted to gratefully acknowledge the Universiti Putra Malaysia for giving me the opportunity to complete my study during my Ph.D journey. Thanks also are due to other members of the academic, and the technical staff in the Faculty of Computer Science and Information Technology for their help and effort to provide facilities, equipments, and an excellent environment to accomplish this research.

My acknowledgement would be incomplete without thanking the biggest source of my strength, my family. The blessings of parents Mrs. Nasrin & Mr. Ahmad and the love and care of my sisters Ghoncheh and Nastaran. I thank them for putting up with me in difficult moments where I felt stumped and for goading me on to follow my dream of getting this degree. This would not have been possible without their unwavering and unselfish love and support given to me at all times.

I would like to dedicate this work to my late mother Mrs. Nasrin Salarzadeh whose dreams for me have resulted in this achievement and without her loving upbringing and nurturing; I would not have been where I am today and what I am today. It is true that if god ever existed, he would be in the form of a mother, because only a mother can love and give without expecting anything in return. Had it not been for my

mother's unflinching insistence and support, my dreams of excelling in education would have remained mere dreams. I thank my mother with all my heart and I know she is up there, listening, watching over me and sending me her blessings constantly and is my guardian angel.

Ghazaleh Babanejad Dehaki

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of PhD. The members of the Supervisory Committee were as follows:

Hamidah Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Nor Izura Udzir, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Fatimah Sidi, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Ali A. Alwan Aljuboori, PhD
Assistant Professor
Faculty of Information and Communication Technology
International Islamic University Malaysia
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 11 February 2021

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: ______________________ Date: ______________

Name and Matric No: Ghazaleh Babanejad Dehaki, GS37718

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Query processing which extracts data items[1] from a database according to a set of access criteria, also known as conditions, and presents these data items to the user for use, has achieved tremendous success at both research and industry levels. There are many types of queries that have been introduced mainly to accommodate the different needs of applications or systems. For instance, a temporal query (based on temporal query language) retrieves time-referenced or temporal data for applications that require information relating to the past, present, and future time. On the other hand, a spatial query which uses geometry data types such as points, lines, and polygons and considers the spatial relationships between these geometries; is useful in Geographic Information System (GIS), Multimedia Information System (MIS), or Computer Aided-Design (CAD).

The traditional query processing operates either by retrieving data items from a database that strictly satisfy each condition specified in the query or returning an empty result if otherwise. The recent developments in query processing attempt to relax these stringent requirements, by retrieving the best, most preferred data items from a database according to the conditions specified in the query, also known as user-defined preferences. These queries known as preference queries employ preference evaluation techniques, have achieved significant success, as they are widely used in applications related to multi-criteria decision support. During the two past decades, several preference evaluation techniques have been introduced, among them are: top-*k* (Surajit Chaudhuri and Luis Gravano, 1999), skyline (Stephan Börzsönyi et al., 2001; Donald Kossmann et al., 2002; Jan Chomicki et al., 2003; Yidong Yuan et al., 2005; Jian Pei et al., 2005; Parke Godfrey et al., 2005; Jan Chomicki et al., 2005; IIaria Bartolini et al., 2006; Man Lung Yiu et al., 2007; Yuan Fang et al., 2010), *k*-dominance (Chee-Yong Chan et al., 2006a), top-*k* dominating (Man Lung Yiu and Nikos Mamoulis, 2009), and *k*-frequency (Chee-Yong Chan et al., 2006b).

Skyline queries rely on the notion of Pareto dominance filter the data items from a database by keeping only those data items that are not worse than any other. It is a well-known technique that is utilised to identify the best, most preferred data items, also known as skylines, from a database to meet the user's preferences. Consider a user who wanted to go for a holiday with the following preferences: (i) hotel that is nearest to the beach (minimum distance) and (ii) hotel with the cheapest price (minimum price). Generally, hotels that are near to a beach are expensive as compared to those which are far away from a beach, which implies that the chances to find a hotel that meets both preferences are nil. Taking this into consideration, the user is left

[1] Without loss of generality, we use the term data item throughout this thesis to be in line with other research works in similar area. The terms *data*, *object*, *record*, and *tuple* can also be used in this context.

with three choices: (i) hotel that is nearest to the beach (minimum distance) while the price is not the cheapest, (ii) hotel with the cheapest price (minimum price) while it is not the nearest hotel to the beach, and (iii) hotel(s) with price cheaper than hotel (i) and distance nearer than (ii). Eventually, the user will have to make the final decision by choosing a hotel from these filtered hotels. Unlike the traditional query which will obviously return an empty result since there is no hotel with minimum price and minimum distance, the skyline query which relies on the powerful skyline operator introduced by Börzsönyi et al. (2001) managed to return results that are not dominated by any other based on the user-defined preferences.

Since the introduction of skyline queries, there are a lot of research works that have been conducted mainly to solve the optimisation problem in computing the skylines (Kian-Lee Tan et al., 2001; Jan Chomicki et al., 2003; Donald Kossmann et al., 2002; Parke Godfrey et al., 2005; Dimitris Papadias et al., 2003; Ilaria Bartolini et al., 2006). The skyline operator introduced by Börzsönyi et al. (2001) only works with the assumption that data items in the database are comparable. However, in today's era, the presence of incomplete data in a database is inevitable. Furthermore, databases need to frequently change their state to reflect the current and latest information of the applications. The incompleteness and dynamism nature of data make the process of identifying skylines no longer a trivial task. This thesis takes the challenge to solve the problem associated to identifying skylines over a dynamic and incomplete database.

1.2 Problem Statement

Skyline query has been studied extensively since the introduction of skyline operator by Borzsonyi et al. in 2001. Since then, a significant number of skyline algorithms have been proposed, mostly attempt to resolve the optimisation problem that is mainly associated with reduction in the processing time of skyline computations. With this regard, most of the studies calculate the number of pairwise comparisons that needs to be performed in filtering the database to keep only those data items that are not worse than any other. Solving the optimisation problem is crucial especially when dealing with huge database with millions of objects and large number of dimensions. Comparing each pair of data items in the database without any optimisation is inefficient. Many variants of skyline algorithms have evolved; among the notable algorithms include *Divide-and-Conquer* (*D&C*), *Block Nested Loop* (*BNL*) (Stephan Börzsönyi et al., 2001), *Bitmap and Index* (Kian-Lee Tan et al., 2001), *Sort Filter Skyline* (*SFS*) (Jan Chomicki et al., 2003), *Nearest Neighbor* (*NN*) (Donald Kossmann et al., 2002), *Linear Elimination Sort Skyline* (*LESS*) (Parke Godfrey et al., 2005), *Branch and Bound Skyline* (*BBS*) (Dimitris Papadias et al., 2003), and *Sort and Limit Skyline algorithm* (*SaLSa*) (Ilaria Bartolini et al., 2006). However, these algorithms are designed with a rigid assumption that the database is complete. Obviously, with this assumption, all data items in the database are comparable.

In the present information age, most real-world applications often deal with data that are partly missing or incomplete. There are many reasons that give rise to the existence of incomplete data in a database. Among them are negligence in data entry, inaccurate data from heterogeneous data sources, and integrating heterogeneous schemas (Garrett Wolf et al., 2009; Mohamed A. Soliman et al., 2010). The incompleteness of data can be viewed as data items that are missing as a whole or dimension values of a data item that are absent, indicated by a null value. In this thesis, we dealt only with the second view, i.e. we cover only missing dimension values. The skyline algorithms proposed for databases with the assumption that they are complete are not suitable for databases with incomplete data. Deriving skylines in incomplete database is not as straightforward as deriving skylines for a complete database. The skyline algorithms in such situation will have to deal with several issues besides the optimisation problem. The missing values in databases give a negative influence on the number of pairwise comparisons that needs to be performed between the data items. Moreover, the transitivity property of skylines is no longer hold. Cyclic dominance is another issue that needs to be tackled as it yields empty skyline results (Mohamed E. Khalefa et al., 2008). To solve the above issues, several algorithms have been proposed which include *ISkyline* (Mohamed E. Khalefa et al., 2008), *SIDS* (Rahul Bharuka et al., 2013), and *Incoskyline* (Ali A. Alwan et al., 2016).

Databases are dynamic in nature in which their states change throughout the time. These changes are necessary as databases must reflect the current and latest information of the applications. The changes are normally achieved through data manipulation operations (like insert, delete, update operations) and data definition operations (like alter table). The skylines derived before changes are made towards the initial database are no longer valid in the new state of the database. Utilising the existing skyline algorithms would require embarking the algorithms on the new state of the database. However, computing the skylines over the entire database after changes are made is inefficient as not all the data items are affected by the changes. Specifically, this thesis attempts to overcome the following two challenges:

Challenge 1: As deliberated in the above section, there are a lot of skyline algorithms that have been proposed (*Divide-and-Conquer* (*D&C*), *Block Nested Loop* (*BNL*), *Bitmap and Index*, *Sort Filter Skyline* (*SFS*), *Nearest Neighbor* (*NN*), *Linear Elimination Sort Skyline* (*LESS*), *Branch and Bound Skyline* (*BBS*), *Sort and Limit Skyline algorithm* (*SaLSa*), etc) that mainly dealt with the optimisation problem with the assumption that the database is complete. A few skyline algorithms have been proposed, namely: *ISkyline* (Mohamed E. Khalefa et al., 2008), *SIDS* (Rahul Bharuka et al., 2013), and *Incoskyline* (Ali A. Alwan et al., 2014) to tackle the issues related to the incompleteness of data in a database. These algorithms are not suitable for a dynamic database as they blindly examining the entire database after changes are made to derive the skylines which is inefficient as not all data items are affected by the changes. Undoubtedly, this incurs unnecessary computations of skylines. Hence, an efficient method is needed to avoid unnecessary skyline computations when changes are made towards the database either thru a data manipulation operations or data definition operations. To achieve a comprehensive solution, the method will also need to consider the possibility of having incomplete data in the database. Hence, besides the main issue of *optimisation*, issues associated to the presence of incomplete data, namely: *transitivity property* and *cyclic dominance*, need to be tackled as well.

Challenge 2: In order to avoid the recomputation of skylines when a new set of skylines needs to be identified, i.e. unnecessary pairwise comparisons between data items when changes are made towards the database (as stated in Challenge 1), it is essential to retain the domination relationships between data items that are identified when pairwise comparisons are performed. The dominance relationships are identified when skylines are derived based on (i) the initial database, (ii) the new state of the database owing to inserting a new data item(s), deleting or updating an existing data items(s), and (iii) the new state of the database owing to adding a new dimension(s) or removing an existing dimension(s). A mechanism that identifies the relevant information to be retained to be utilised later in the process of identifying a new set of skylines with the aim at avoiding unnecessary computations of skylines is needed. Keeping track of each dominance relationship is unwise as not only it will incur unnecessary storage cost, also not all the dominance relationships will be utilised in the subsequent processes of skyline computations. Hence, besides the main issue of *optimsation*, identifying the *prominent* dominance relationships among all possible dominance relationships is another issue to be dealt with.

1.3 Objective of the Research

The main goal of this research work is to propose an efficient skyline computation framework that is able to process skyline queries over a dynamic and incomplete database. To achieve this goal, the following objectives are set:

(i) To propose an efficient framework that avoids unnecessary skyline computations when changes are made towards an incomplete database due to a data manipulation operation(s).
(ii) To propose an efficient framework that avoids unnecessary skyline computations when changes are made towards an incomplete database due to a data definition operation(s).
(iii) To propose an approach that identifies the relevant information to be retained, to be utilised later in the process of identifying a new set of skylines with the aim of avoiding unnecessary computations of skylines.

1.4 Research Scope

The scope of this research work is outlined as follows:

(i) This research assumed a relational database model as it is the most dominant model that is widely used by almost all businesses and is reflected in the major software offerings from Oracle, SQL Server, etc (Yidong Yuan et al., 2005; Xuemin Lin et al., 2007; Dalie Sun et al., 2008; Mohamed E. Khalefa et al., 2008; Garrett Wolf et al., 2009; Justin J. Levandoski et al., 2010; Ken C. K. Lee et al., 2010).

(ii) The incompleteness of data can be viewed as data items that are missing as a whole or dimension values of a data item that are absent, indicated by a null value. In this thesis, we dealt only with the second view, i.e. we cover only missing dimension values. This is in line with other works that have contributed skyline algorithms over an incomplete database (Mohamed E. Khalefa et al., 2008; Rahul Bharuka et al., 2013; Ali A. Alwan et al., 2014).

(iii) The state of the database changes throughout its lifetime to reflect the current and latest information of the applications. The changes are normally achieved either through data manipulation operations (like insert, delete, update operations) or data definition operations (like alter table). This thesis considers both type of operations although changes due to data definition operations are infrequent as compared to data manipulation operations.

1.5 Organisation of the Thesis

This thesis is organised as follows:

Chapter 1 is the introduction chapter of the thesis which starts with an overall overview and the motivation behind this study. The problem statement, the research objectives as well as the research scope are deliberated in this chapter.

Chapter 2 is the literature review chapter that gives a brief overview of preference queries in database systems with focus given mainly on skyline queries. There is a large volume of published works on preference queries which are reviewed in this chapter. Discussions include the Top-*k*, Top-*k* dominating, *K*-dominance, *K*-frequency, and skyline preference queries. In this chapter the features, strength, and weaknesses of these preference queries are highlighted.

Chapter 3 presents the research methodology of the research reported in this thesis. This chapter begins with the theoretical dimensions of the research and presents the phases that are conducted in achieving the main goal of this research work. It also presents the performance measurement metrics and the data sets that are utilised in the experiments of this research work.

Chapter 4 describes the design and gives the detail phases and algorithms of the proposed framework for processing skyline queries over an incomplete database. The chapter deliberates on the proposed approach in identifying prominent dominance relationships to be saved and utilised later by the subsequent processes. The phases of the proposed framework are illustrated step by step with a running example.

Chapter 5 describes the design and gives the detail phases and algorithms of the proposed framework for processing skyline queries when a database is changed due to a data manipulation operation(s), i.e. insert a new data item(s), delete or update an existing data item(s). In this chapter we show how the saved prominent dominance relationships identified in Chapter 4 are utilised with the aim to avoid unnecessary computations of skylines. The phases of the proposed framework are illustrated step by step with a running example.

Chapter 6 describes the design and gives the detail phases and algorithms of the proposed framework for processing skyline queries when a database is changed due to a data definition operation, i.e. add a new dimension(s) and remove an existing dimension(s). Similar to chapter 5, the saved prominent dominance relationships identified in Chapter 4 are utilised with the aim to avoid unnecessary computations of skylines. The phases of the proposed framework are illustrated step by step with a running example.

Chapter 7 presents the experiments that have been carefully designed and conducted in order to accurately evaluate the performance of the proposed frameworks. The results of the experiments are reported and these results are compared to the previous works that are related to the study to ascertain the improvement gained. The experiments are evaluated with different parameters that are data set size, number of dimensions, number of dimensions with missing values, and changing rate.

Chapter 8 presents the conclusion and contributions of the research works. Some recommendations for the future works are also listed.

REFERENCES

Adegbemiga Ola and Gultekin Ozsoyoglu. 1993. A Family of Incomplete Relational Database Models. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, No. 2, pp. 293-308.

Akrivi Vlachou, Christos Doulkeridis, Kjetil Nørvåg, and Michalis Vazirgiannis. 2008. On Efficient Top-k Query Processing in Highly Distributed Environments, *Proceedings of the International Conference on Management of Data (ICMD08)*, Vancouver (Canada), pp. 753-764.

Alan C. Acock. 2005. Working with Missing Values. *Journal of Marriage and Family*, Vol. 67, No. 4, pp. 1012–1028.

Ali A. Alwan, H. Ibrahim, N. I. Udzir, and F. Sidi. 2016."An Efficient Approach for Processing Skyline Queries in Incomplete Multidimensional Database," *Arabian Journal for Science and Engineering*, vol. 41, pp. 2927-2943.

Alon Y. Levy. 1996. Obtaining Complete Answers from Incomplete Databases, *Proceedings of the 22nd International Conference on Very Large Data Base*. Mumbai (Bombay), India, pp. 402-412.

Anastasios Arvanitis and Georgia Koutrika. 2012. Towards Preference-aware Relational Databases, *Proceedings of the International Conference on Data Engineering, (ICDE2012)*, Washington DC (USA), pp. 426 – 437.

Anish Das Sarma, Omar Benjelloun, Alon Y. Halevy, and Jennifer Widom. 2006. Working Models for Uncertain Data. *Proceedings of the 22rd International Conference on Data Engineering, (ICDE 2006)*, Atlanta, Georgia, (USA), pp. 7.

Beng Chin Ooi, Cheng Hian Goh, and Kian-Lee Tan. 1998. Fast High-Dimensional Data Search in Incomplete Databases, *Proceedings of the 24th International Conference on Very Large Data Base, VLDB 1998*, New York (USA), pp. 357 – 367.

Bhekisipho Twala, Michelle Cartwright, and Martin J. Shepperd. 2005. Comparison of Various Methods for handling Incomplete Data in Software Engineering Databases, *Proceedings of the International Symposium on Empirical Software Engineering. Noosa Heads (Australia)*, pp. 105-114.

Bin Jiang, Jian Pei, Xuemin Lin, and Yidong Yuan. 2012. Probabilistic Skylines on Uncertain Data: Model and Bounding-Pruning-Refining Methods. *Journal of Intelligent Information Systems*, Vol. 38, No. 1, pp 1-39.

Brian Babcock and Chris Olston: Distributed Top-k Monitoring. 2003, *Proceedings of the International Conference on Management of Data*, San Diego, California, (USA), pp. 28-39.

Chee-Yong Chan, H. V. Jagadish, Kian-Lee Tan, Anthony KH Tung, and Zhenjie Zhang. 2006. Finding k-dominant skylines in high dimensional space. *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data*, pp. 503-514.

Chee-Yong Chan, Jagadish, H.V., Kian-Lee, T., Anthony K.H. T., and Zhenjie, Z. 2006(a). On High Dimensional Skylines, *Proceedings of the 10th International Conference on Extending Database Technology, EDBT'06*, Munich (Germany), pp. 478-495.

Chee-Yong Chan, Jagadish, H.V., Kian-Lee, T., Anthony K.H. T., and Zhenjie, Z. 2006(b). Finding K-dominant Skylines in High Dimensional Space, *Proceedings of the International Conference on Management of Data (ICMD06)*, Chicago, Illinois (USA), pp. 503-514.

Cheng Luo, Zhewei Jiang, Wen-Chi Hou, Shan He, and Qiang Zhu. 2012. A Sampling Approach for Skyline Query Cardinality Estimation. *Knowledge and Information Systems*, Vol. 32, No. 2, pp 281-301.

CV Neethu and Rejimol Robinson. 2013. A Survey of Techniques For Answering Top-k queries. *Global Journal of Computer Science and Technology*.

Dalie Sun, Sai Wu, Jianzhong Li, and Anthony K. H. Tung. 2008. Skyline-Join in Distributed Databases, *Proceedings of the 24th International Conference of Data Engineering Workshop (ICDE Workshop08)*, Cancun (Mexico), pp. 176-181.

Dimitris Papadias, Yufei Tao, Greg Fu, and Bernhard Seeger. 2005. Progressive skyline computation in database systems. *ACM Transactions on Database Systems*, Vol. 30, No. 1, pp. 41–82.

Dimitris Sacharidis, Panagiotis Bouros, and Timos Sellis. 2008. Caching dynamic skyline queries. *Proceedings of the International Conference on Scientific and Statistical Database Management Springer*, Berlin, Heidelberg. pp. 455-472.

Dimitrios Skoutas, Dimitris Sacharidis, Alkis Simitsis, Verena Kantere, and Timos Sellis. 2009. Top-k dominant web services under multi-criteria matching, *Proceedings of the 12th international conference on extending database technology: advances in database technology* (*ACM*). (pp. 898-909).

Donald Kossmann, Frank Ramsak, and Steffen Rost. 2002. Shooting Stars in the Sky: An Online Algorithm for Skyline Queries, *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB 28)*, Hong Kong (China), pp. 275- 286.

Garrett Wolf, Aravind Kalavagattu, Hemal Khatri, Raju Balakrishnan, Bhaumik Chokshi, Jianchun Fan, Yi Chen, and Subbarao Kambhampati. 2009. Query Processing Over Incomplete Autonomous Databases: Query Rewriting Using Learned Data Dependencies. *The International Journal on Very Large Data Bases*, Vol.18, No. 5, pp. 1167 – 1190.

Gerard Salton, and Michael J. McGill. 1983. Introduction to modern information retrieval. *mcgraw-hill*.

Guadalupe Canahuate, Michael Gibas, and Hakan Ferhatosmanoglu. 2006. Indexing Incomplete Databases, *Proceedings of the 10th International Conference on Advances in Database Technology (EDBT)*, Munich, Germany, pp. 884-901.

Gustavo Batista and Maria Carolina Monard. 2003. An Analysis of Four Missing Data Treatment Methods for Supervised Learning. *Applied Artificial Intelligence Journal*, Vol. 17, pp. 519-533.

Haoyang Zhu, Peidong Zhu, Xiaoyong Li and Qiang Liu. 2017. Top-k Skyline Groups Queries. *Proceedings of the 20th International Conference on Extending Database Technology (EDBT)*, Venice, Italy, pp. 442-445

Hyunsik Choi, Harim Jung, Ki Yong Lee, and Yon Dohn Chung. 2013. Skyline queries on keyword-matched data. *Information Sciences 232*, pp. 449-463.

I-Fang Su, Yu-Chi Chung, and Chiang Lee. 2010. Top-k combinatorial skyline queries. *Proceedings of the International Conference on Database Systems for Advanced Applications*, Berlin, Heidelberg, pp. 79-93.

Ihab F. Ilyas, Walid G. Aref, and Ahmed K. Elmagarmid. 2003. Supporting Top-k Join Queries in Relational Databases, *Proceedings of the 29th International Conference on Very Large Data Bases (VLDB 29)*, Berlin (Germany), pp. 754-765.

Ihab F. Ilyas, Walid G. Aref, and Ahmed K. Elmagarmid. 2004. Supporting Top-k Join Queries in Relational Databases. *Very large Database Journal VLDB*, Vol.13 No.3, pp. 207-221.

Ihab F. Ilyas, George Beskales, and Mohamed A. Soliman. 2008. A survey of top-k query processing techniques in relational database systems. *ACM Computing Surveys (CSUR)*, pp. 1-58.

Ilaria Bartolini, Paolo Ciaccia, and Marco Patella. 2006. SaLSa: Computing the Skyline without Scanning the Whole Sky, *Proceedings of the 15th International Conference on Information and Knowledge Management (ICIKM06)*, USA, Arlington, pp. 405- 414.

Jan Chomicki, Parke Godfrey, Jarek Gryz, and Dongming Liang. 2003. Skyline with Presorting, *Proceedings of the 19th International Conference on Data Engineering (ICDE03)*, Bangalore, India, pp.717-816.

Jan Chomicki, Parke Godfrey, Jarek Gryz, and Dongming Liang. 2005. Skyline with Presorting: *Theory and Optimizations. Intelligent Information Systems Journal*, Vol. 31, pp. 595-604.

Jerzy W. Grzymala-Busse and Ming Hu. 2000. A Comparison of Several Approaches to Missing Attribute Values in Data Mining, *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing*, Canada, Banff, pp. 378 – 385.

Jerzy W. Grzymala-Busse. 2004(a). Rough Set Approach to Incomplete Data. *Proceedings of the 7th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2004)*, Poland, Zakopane, pp. 50-55.

Jerzy W. Grzymala-Busse. 2004(b). Data with Missing Attribute Values: Generalization of Indiscernibility Relation and Rule Induction. *Transactions on Rough Sets*, pp. 78-95.

Jerzy W. Grzymała-Busse and Wojciech Rza sa. 2006. Local and Global Approximations for Incomplete Data. *Rough Sets and Current Trends in Computing Lecture Notes in Computer Science*, Vol. 4259, pp 244-253.

Jerzy W. Grzymala-Busse and Wojciech Rzasa. 2008. Local and Global Approximations for Incomplete Data. *Transactions on Rough Sets*, Vol. 8, pp. 21-34.

Jian Pei, Wen Jin, Martin Ester, and Yufei Tao. 2005. Catching the Best Views of Skyline: A Semantic Approach Based on Decisive Subspaces, *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 31)*, Norway, Trondheim, pp. 253-264.

Jongwuk Lee, Gae-won, Y., and Seung-won, H. 2009. Personalized Top-k Skyline Queries in High-Dimensional Space. *Information Systems Journal*, Vol. 34, No. 1, pp. 45-61.

Jongwuk Lee, Hyeonseung Im, and Gae-won You. 2016."Optimizing Skyline Queries over Incomplete Data," *Information Sciences*, vol. 361, pp. 14-28, 2016.

Justin J. Levandoski, Mohamed F. Mokbel, and Mohamed E. Khalefa. 2010. FlexPref: A Framework for Extensible Preference Evaluation in Database Systems, *Proceedings of the 26th International Conference on Data Engineering, (ICDE2010)*, USA, Long Beach, California, pp. 828-839.

Kaiqi Zhang, Hong Gao, Xixian Han, Zhipeng Cai, and Jianzhong Li. 2016. ISSA: Efficient skyline computation for incomplete data. *Proceedings of the International Conference on Database Systems for Advanced Applications Springer*, Cham, pp. 321-328.

Kaiqi Zhang, Hong Gao, Xixian Han, Zhipeng Cai, and Jianzhong Li. 2017. Probabilistic skyline on incomplete data. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 427-436.

Ken C. K. Lee, W., Zheng, B., Li, H., and Tian Y. 2010. Z-SKY: An Efficient Skyline Query Processing Framework Based on Z-order. *Journal of Very Large Database*, *VLDB*, Vol. 19, No. 3, pp. 333-362.

Keping Zhao, Yufei Tao, and Shuigeng Zhou. 2007. Efficient Top-k Processing in Large-scaled Distributed Environments. *Data and Knowledge Engineering Journal* Vol. 63, No. 2, pp. 315-335.

Kevin Chen-chuan, Chang and Seung-won Hwang. 2002. Minimal Probing: Supporting Expensive Predicates for Top-k Queries, *Proceedings of the International Conference on Management of Data (ICMD02)*, Madison, Wisconsin (USA), pp. 346 - 357.

Kian-Lee Tan, Pin-Kwang, E., and Beng, C O. 2001. Efficient Progressive Skyline Computation, *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB27)*, Roma (Italy), pp. 301-310.

Ke Yi, Hai Yu, Jun Yang, Gangqiang Xia, and Yuguo Chen. 2003. Efficient Maintenance of Materialized Top-k Views, *Proceedings of the 19th International Conference on Data Engineering (ICDE 2003)*, Bangalore, (India), pp. 189-200.

Kyriakos Mouratidis, Spiridon Bakiras, and Dimitris Papadias. 2006. Continuous Monitoring of Top-k Queries over Sliding Windows, *Proceedings of the International Conference on Management of Data (ICMD06)*, Chicago, Illinois (USA), pp. 635-646.

Lei-gang Dong, Xiao-wei Cui, Zhen-fu Wang, Ying-rui Ma, and Guo-qiang Shao. 2011. Updating skyline with dynamic space set. *Proceedings of the Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 3, pp. 1531-1534.

Leonid Libkin. 2006. Data exchange and incomplete information. *Proceedings of the 25th ACM International Symposium on Principles of Database Systems (PODS2006)*, Chicago, Illinois, (USA), pp. 60-69.

Lyublena Antova, Christoph Koch, and Dan Olteanu 2009. 10(10)6 Worlds and Beyond: Efficient Representation and Processing of Incomplete Information. *The Very Large Database Journal VLDB*, Vol.18, pp. 1021–1040.

Lyublena Antova, Christoph Koch, and Dan Olteanu. 2007. From Complete to Incomplete Information and Back, Proceedings of the 2007 ACM SIGMOD *International Conference on Management of Data, SIGMOD'07*, Beijing, (China), pp. 713 – 724.

Man Lung Yiu and Nikos Mamoulis. 2007. Efficient Processing of Top-k Dominating Queries on Multi-dimensional Data, *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB32)*, Trondheim (Norway), pp. 483-494.

Man Lung Yiu and Nikos Mamoulis. 2009. Multi-dimensional Top-k Dominating Queries. *The Very Large Database Journal VLDB*, Vol. 18, No. 3, pp. 695-718.

Maria Kontaki, Apostolos N. Papadopoulos, and Yannis Manolopoulos. 2008. Continuous Top-k Dominating Queries in Subspaces, *Proceedings of the 12th Panhellenic Conference on Informatics (PCI 2008)*, Samos Island, (Greece) pp. 31 – 35.

Maria Kontaki, Apostolos N. Papadopoulos, and Yannis Manolopoulos. 2010. Continuous Processing of Preference Queries in Data Streams, *Proceedings of the 36th International Conference on Current Trends in Theory and Practice of Computer Science, Špindleruv Mlýn*, Czech Republic, pp. 47-60.

Maria Kontaki, Apostolos N. Papadopoulos, and Yannis Manolopoulos. 2011. Continuous top-k dominating queries. *IEEE Transactions on Knowledge and Data Engineering 24*, no. 5: pp. 840-853.

Martin Theobald, Gerhard Weikum, and Ralf Schenkel. 2004. Top-k Query Evaluation with Probabilistic Guarantees, *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB 30)*, Toronto (Canada), pp. 648-659.

Md Anisuzzaman Siddique, and Yasuhiko Morimoto. 2010. K-Dominant and Extended k-dominant Skyline Computation by using Statistics. *International Journal on Computer Science and Engineering*, Vol. 02, No.05, pp. 1934-1943.

Michael D. Morse, Jignesh M. Patel, and William I. Grosky. 2007. Efficient Continuous Skyline Computation, *Information Science Journal*, Vol. 177, No. 17, pp. 3411-3437.

Ming Zhang and Reda Alhajj. 2010. Skyline Queries with Constraints: Integrating Skyline and Traditional Query Operators. *Data & Knowledge Engineering Journal*, Vol. 69, No. 1, pp. 153- 168.

Mohamed E. Khalefa, Mohamed F. Mokbel and Justin J. Livandoski. 2008. Skyline Query Processing For Incomplete Data, *Proceedings of the 24th International Conference on Data Engineering (ICDE 2008)*, Cancun, (Mexico), pp. 556-565.

Mohamed A. Soliman, Ihab F. Ilyas., Shalev, and Ben-David. 2010. Supporting Ranking Queries on Uncertain and Incomplete data. *The Very Large Database Journal, VLDB*, Vol. 19, No. 4, pp. 477- 501.

Mohamed A. Soliman, Ihab F. Ilyas, and Kevin Chen-Chuan Chang. 2007. Top-k Query Processing in Uncertain Databases, *Proceedings of the 23rd International Conference on Data Engineering (ICDE 2007)*, Istanbul (Turkey), pp. 896 – 905.

Mohammad Shamsul Arefin, and Morimoto, Y. 2012. Skyline sets queries for incomplete data. *International Journal of Computer Science & Information Technology*, *4*(5), pp. 67.

Nicolas Bruno, Surajit Chaudhuri, and Luis Gravano. 2002. Top-k Selection Queries over Relational Databases: Mapping Strategies and Performance Evaluation. *ACM Transactions on Database Systems*, Vol. 27, No. 2, pp. 153–187.

Parisa Haghani, Sebastian Michel, and Karl Aberer. 2009. Evaluating top-k queries over incomplete data streams. *Proceedings of the 18th International Conference on Information and Knowledge Management (CIKM2009)*, Hong Kong, (China), pp. 877-886.

Parke Godfrey. 2004. Skyline Cardinality for Relational Processing. Foundations of Information and Knowledge Systems. *Lecture Notes in Computer Science*, Vol. 2942, pp. 78-97, 2004.

Parke Godfrey, Ryan Shipley, and Jarek Gryz. 2005. Maximal Vector Computation in Large Data Sets, *Proceedings of the 31st International Conference on Very Large Data* Bases (VLDB30), Toronto, Ontario, (Canada), pp. 229-240.

Per Jonsson and Claes Wohlin. 2004. An Evaluation of k-Nearest Neighbour Imputation Using Likert Data. *Proceedings of the 10th International Symposium on Software Metrics*, Chicago, IL, (USA), pp. 108 – 118.

Ping Wu, Divyakant Agrawal, Omer Egecioglu, and Amr El Abbadi. 2007. Deltasky: Optimal maintenance of skyline deletions without exclusive dominance region generation. *Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering*, pp. 486-495.

Rahul Bharuka, and P. Sreenivasa Kumar. 2013. Finding skylines for incomplete data. Proceedings of the Twenty-Fourth *Australasian Database Conference, Australia*, Vol. 137, pp. 109-117.

Raymond Chie-Wing Wong, Ada Wai-Chee Fu, Jian Pei, Yip Sing Ho, Tai Wong, and Yubao Liu. 2008. Efficient Skyline Querying With Variable User Preferences on Nominal Attributes, *Proceedings of the 34th International Conference on Very Large Data Bases (VLDB 34)*, Auckland, (New Zealand), pp. 1032-1043.

Reza Akbarinia, Esther Pacitti, and Patrick Valduriez. 2007. Processing Top-k Queries in Distributed Hash Tables. *Proceedings of the 13th International Euro-Par Conference*, Rennes, (France), pp. 489-502.

Sebastian Michel, Peter Triantafillou, and Gerhard Weikum. 2005. KLEE: A Framework for Distributed Top-k Query Algorithms. *Proceedings of the 31st International Conference on Very Large Data Base, (VLDB31)*, Trondheim, (Norway), pp. 637-648.

Shashi K. Gadia, Sunil S. Nair, and Yiu-Cheong Poon. 1992. Incomplete Information in Relational Temporal Databases. *Proceedings of the 18th International Conference on Very Large Data Bases (VLDB 18)*, Vancouver, British Columbia, (Canada), pp. 395-406.

Sheldon Shen. 1988. Database Relaxation: An Approach to Query Processing in Incomplete Databases. *Information Processing and Management Journal*, Vol. 24, No. 2, pp. 151 – 159.

Shingo Otsuka and Nobuyoshi Miyazaki. 1998. An Incomplete Database Approach to Global Query Processing, *Proceedings of the 13th International Conference on Information Networking ICOIN '98*, Koganei, Tokyo (Japan), pp. 337 – 342.

Simon Razniewski and Werner Nutt. 2011. Completeness of Queries over Incomplete Databases, *Proceedings of the 37th International Conference on Very Large Data Base, (VLDB37)*, Seattle, Washington (USA), pp. 749-760.

Stephan Börzsönyi, Donald Kossmann, and Konrad Stocker. 2001. The Skyline Operator, *Proceedings of the 17th International Conference on Data Engineering (ICDE01)*, Cancun, (Mexico),pp. 421-430.

Surajit Chaudhuri and Luis Gravano 1999. Evaluating Top-k Selection Queries. *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB 25)*, Edinburgh, (Scotland), 7-10 September, 397-410.

Surajit Chaudhuri, Luis Gravano, and Amélie Marian. 2004. Optimizing Top-k Selection Queries over Multimedia Repositories. *IEEE Transaction Knowledge and Data Engineering*, Vol. 16, No. 8, pp. 992-1009.

Suvarna Bothe, Panagiotis Karras, and Akrivi Vlachou. 2013. eskyline: Processing skyline queries over encrypted data. *Proceedings of the VLDB Endowment 6*, no. 12: 1338-1341.

Tassadit Bouadi, Cordier, M. O., & Quiniou, R. 2012. Incremental computation of skyline queries with dynamic preferences. *Proceedings of the International Conference on Database and Expert Systems Applications* (pp. 219-233). Springer, Berlin, Heidelberg.

Todd J. Green and Val Tannen. 2006. Models for Incomplete and Probabilistic Information. *IEEE Data Engineering Bulletin*, Vol. 29, No. 1, pp. 17-24.

Tomasz Imieliski and Witold Lipski. 1984. Incomplete Information in Relational Databases. *Journal of the Association for Computing Machinery*, Vol. 31, No 4, pp. 761-791.

Vagelis Hristidis and Yannis Papakonstantinou. 2004. Algorithms and Applications for Answering Ranked Queries Using Ranked Views. *The Very Large Data Base Journal, VLDB*, vol. 13, No. 1, pp. 49-70, 2004.

Veronique Bruyere, Alexandre Decan and Jef Wijsen. 2009. On First-Order Query Rewriting for Incomplete Database Histories, *Proceedings of the 2009 16th International Symposium on Temporal Representation and Reasoning*, Bressanone-Brixen (Italy), pp. 54-61.

Vineetha Sara Abraham, and S. Deepa Kanmani. 2013. A Survey on Continuous Monitoring of Preference Queries Using Sliding Window. *International Journal of Scientific and Research Publications*, Vol 3.

Wanbin Son, Fabian Stehn, Christian Knauer, and Hee-Kap Ahn. 2017. Top-k manhattan spatial skyline queries. *Information Processing Letters 123*: pp. 27-35.

Xiaofeng Ding, and Hai Jin. 2012.Efficient and progressive algorithms for distributed skyline queries over uncertain data. *IEEE Transactions on Knowledge and Data Engineering 24*, no. 8 pp. 1448-1462.

Xiaoye Miao, Yunjun Gao, Gang Chen, & Huiyong Cui. 2015. Top-k dominating queries on incomplete data. *IEEE Transactions on Knowledge and Data Engineering*, 28(1), pp. 252-266.

Xiaoye Miao, Yunjun Gao, Gang Chen, and Tianyi Zhang. 2016. k-dominant skyline queries on incomplete data. *Information Sciences 367*: pp. 990-1011.

Xin Lin, Jianliang Xu, and Haibo Hu. 2014. Range-based skyline queries in mobile environments. *IEEE Transactions on Knowledge and Data Engineering 25*, no. 4 (2011): 835-849.

Xintao Wu and Daniel Barbara. 2002. Learning Missing Values from Summary Constraints. *ACM SIGKDD Explorations Newsletter*, Vol. 4, No.1, pp. 21-30.

Yang C. Yuan. 2000. Multiple Imputation for Missing Data: Concepts and New Development. *Proceedings of the 25th Annual SAS Users Group International Conference*. Indianapolis, Indiana, (USA).

Yan Wang, Zhan Shi, Junlu Wang, Lingfeng Sun, and Baoyan Song. 2017. Skyline preference query based on massive and incomplete dataset. *IEEE Access 5*: pp. 3183-3192.

Yidong Yuan, Xuemin Lin, Qing Liu, Wei Wang, Jeffrey Xu Yu, and Qing Zhang. 2005. Efficient Computation of the Skyline Cube. *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 31)*, Trondheim, (Norway), pp. 267-278.

Ying-Yuan Xiao and Chen Yue-Guo. 2010. Efficient distributed skyline queries for mobile applications. *Journal of Computer Science and Technology*, 25(3), pp. 523-536.

Yonis Gulzar, Ali A. Alwan, Norsaremah Salleh, Imad Fakhri Al Shaikhli, and Syed Idrees Mairaj Alvi. 2016. A framework for evaluating skyline queries over incomplete data. *Procedia Computer Science*, 94, pp. 191-198.

Yuan-Chi Chang, Lawrence Bergman, Vittorio Castelli, Chung-Sheng Li, Ming-Ling Lo, and John R. Smith. 2000. The Onion Technique: Indexing for Linear Optimization Queries, *Proceedings of the International Conference on Management of Data*, Dallas, Texas (USA), pp. 391-402.

Zhenhua Huang and Wei Wang. 2006. A Novel Incremental Maintenance Algorithm of SkyCube, *Proceedings of the 17th International Conference of Database and Expert Systems Applications (DEXA 2006)*, Kraków (Poland), pp. 781-790.

Zhenhua Huang, Shengli Sun, and Wei Wang. 2010. Efficient Mining of Skyline Objects in Subspaces over Data Streams. *Knowledge and Information Systems Journal*, Vol. 22, No. 2, pp. 159-183.

Zhiyong Huang, Hua Lu, Beng Chin Ooi, and Anthony KH Tung. 2006. Continuous skyline queries for moving objects. *IEEE transactions on knowledge and data engineering*, 18(12), pp. 1645-1658.

BIODATA OF STUDENT

Ghazaleh Babanejad Dehaki was born on March 21, 1981 in Tehran, Iran. She received her bachelor in the field of software engineering from faculty of computer science from Iran University of Science and Technology in 2007.

In 2010, she pursued her study for Master of Knowledge Management with Multimedia at Multimedia University (MMU) in Malaysia, specialized in semantic web ontology and recommender systems. She did her research on creating CRM for natural disaster management system using Protégé and semantic web.

In 2013, she continued her study for Ph.D. degree at University Putra Malaysia (UPM) in the field of Database Systems, specialized in Preference Query in Dynamic and Incomplete Database Systems. She did her research on the Efficient Computation of Skyline Queries over a Dynamic and Incomplete Database.

Her interests include Preference Queries, Dynamic and Incomplete Databases, Distributed Database, Recommendation and Decision Support Systems, Big Data and Big Data Analytics.

LIST OF PUBLICATIONS

Conferences

Babanejad, G., Ibrahim, H., Udzir, N. I., Sidi, F., & Aljuboori, A. A. A. (2014). Finding skyline points over dynamic incomplete database. *In Proceedings of Malaysian National Conference on Databases (MaNCoD).*

Babanejad, G., Ibrahim, H., Udzir, N. I., SIDI, F., BABANEJAD, G., & ALWAN, A. (2015). Identifying Skylines in Dynamic Incomplete Database. In Proceedings of the 9th International Conference on Computer Engineering and Applications (CEA15) (pp. 231-236).

Babanejad, G., Ibrahim, H., Udzir, N. Z., Sidi, F., & Alwan, A. A. (2017, April). Deriving skyline points over dynamic and incomplete databases. *In Proceedings of the 6th International Conference of Computing and Informatics,* April (pp. 25-27).

Babanejad, G., Ibrahim, H.,Udzir, N. I., Sidi, F., & Alwan, A. A. (2018, November). Efficient Skyline Processing Algorithm over Dynamic and Incomplete Database. *In Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services* (pp. 190-199).

Babanejad, G., Ibrahim, H., Udzir, N. I., Sidi, F., & Alwan, A. A. (2018). Processing Skyline Algorithms over Dynamic and Incomplete Databases. *In Proceedings of the 29th International Conference on Database and Expert Systems* (DEXA 2018). **Accepted Paper**

UNIVERSITI PUTRA MALAYSIA

STATUS CONFIRMATION FOR THESIS / PROJECT REPORT AND COPYRIGHT

ACADEMIC SESSION : ________________

TITLE OF THESIS / PROJECT REPORT :

____

____

____

NAME OF STUDENT : ______________________________

I acknowledge that the copyright and other intellectual property in the thesis/project report belonged to Universiti Putra Malaysia and I agree to allow this thesis/project report to be placed at the library under the following terms:

1. This thesis/project report is the property of Universiti Putra Malaysia.

2. The library of Universiti Putra Malaysia has the right to make copies for educational purposes only.

3. The library of Universiti Putra Malaysia is allowed to make copies of this thesis for academic exchange.

I declare that this thesis is classified as :

*Please tick (√)

☐	CONFIDENTIAL	(Contain confidential information under Official Secret Act 1972).
☐	RESTRICTED	(Contains restricted information as specified by the organization/institution where research was done).
☐	OPEN ACCESS	I agree that my thesis/project report to be published as hard copy or online open access.

This thesis is submitted for :

☐	PATENT	Embargo from_____________ until ______________ (date) (date)

Approved by:

______________________ (Signature of Student) New IC No/ Passport No.:	______________________ (Signature of Chairman of Supervisory Committee) Name:
Date :	Date :

[Note : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization/institution with period and reasons for confidentially or restricted.]