RESEARCHERID
THOMSON REUTERS

ISSN: 2277-9655
[Saravanapriya* *et al.*, 6(9): September, 2017] Impact Factor: 4.116
IC™ Value: 3.00 CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A SURVEY ON SKYLINE COMPUTING FOR UNCERTAIN DATABASE

S.Saravanapriya [1], Dr.V.Thiagarasu [2]
Research Scholar, Department of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Erode, India.
Associate Professor of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Erode, India.

DOI: Will get Assigned by IJESRT Team

ABSTRACT

Skyline filters out hard and fast of thrilling points from a probably large set of statistics points. A point is thrilling if it is not dominated by using explicit. The skyline queries are vital on the way to assist users to deal with the large amount of available records through figuring out a fixed of interesting records items. Skyline computation is extensively used in multi-standards decision making. This paper conducts a survey on research issues on computing skyline for uncertain databases, with the view of providing interested researchers with an overview of the most recent research directions in this area. It further suggests possible research direction on skyline processing for uncertain databases.

KEYWORDS: Skyline operator, Big data management, Uncertain databases, Big data.

I. INTRODUCTION

1.Big data

Big data is a term for data sets which can be so massive or complicated that conventional data processing application software is inadequate to deal with them proposed by crawford[2011]. Big data is not a single technology but a combination of old and new technologies that helps companies gain actionable insights. Consequently, big information is the capability to control a massive extent of disparate information, at the proper pace, and inside the proper time frame to allow real-time evaluation and response.

Figure1 illustrates that data must first be captured, and then organized and integrated. After this point is strongly implemented, data can be analyzed based on the problem being addressed. Finally, management takes action based on the result of that analysis proposed by A.Wiley brand [2013].

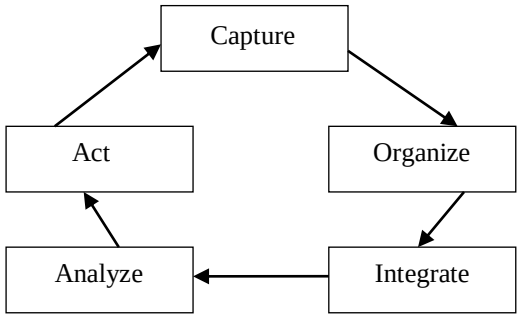

Figure 1: The cycle of big data management

Big data characteristics

Data Volume: The quantity of generated and saved facts. The scale of the information determines the value and ability perception and whether it colud surely be considered large records or no longer.
Data Velocity: In this context, the rate at which the statistics is generated and processed to satisfy the demands and demanding situations that lie inside the path of growth and improvement proposed by Hilbert [2015].
Data Variety: Data comes in all types of formats from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock material data and financial transactions.

Data Value: The amount of generated and stored statistics. The dimensions of the statistics determines the value and capacity insight and whether it can really be taken into consideration massive facts or not.
Complexity: Current data comes from multiple sources, which makes it difficult to link, match, restore and transform data across systems. However, it's necessary to connect and compare relationships, hierarchies and multiple data connection or your data can quickly circling out of control.
Figure 2 illustrates that Variety represents the types of records in data, velocity refers to the rate at which the specific amount of data is generated and analyzed, and volume defines the amount or number of records of data. Veracity means how much amount of the data can be verified given the accuracy of its source proposed by Walunj Swapnil.K [2016].

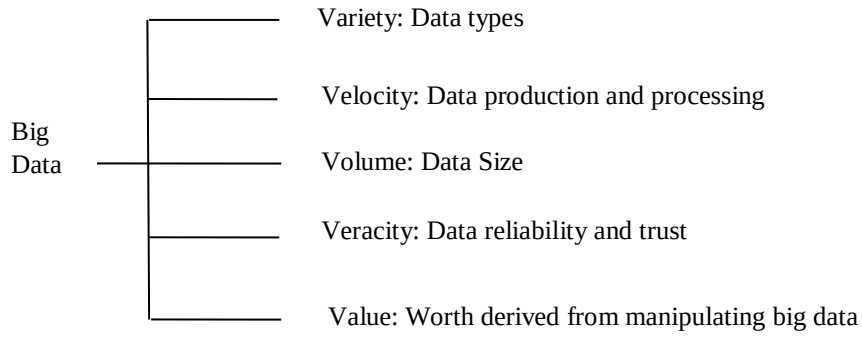

Figure 2: Big data characteristics

2.Skyline operator

Skyline operation filters out a interesting factors from a probably large set of data points. A point is exciting if it is not ruled by way of every other point proposed by Borzsonyi [2001].

3.Uncertain Data

Uncertain data is the data that consists of noise that makes it deviate from the correct, unique values. Inside the age of big data, uncertainty or facts veracity is one of the defining characteristics of data. Data is continuously growing in extent, variety, pace and uncertainty. Uncertain information is found in abundance nowadays on the net, in sensor networks, inside establishments both in their dependent and unstructured resources. As an instance, there can be uncertainty concerning the deal with of a customer in an organisation dataset, or the temperature readings captured through a sensor due to getting old of the sensor proposed by William R. Lfontaine [2012].

4.Big Data Management

Big data management is the corporation, management and governance of big volumes of both established and unstructured information. The aim of large facts control is to ensure a high degree of facts exceptional and accessibility for commercial enterprise intelligence and big data analytic packages. Businesses, government organizations and other organizations hire big data management strategies to help them cope with speedy-developing swimming pools of information, usually involving many terabytes or even petabytes of information stored in a spread of report formats. Effective big data management facilitates agencies locate valuable facts in massive sets of unstructured data and semi-established records from a ramification of assets, which include call element data, machine logs and social media web sites proposed by Melanie Luna[2013].

II. LITERATURE SURVEY

Dongwon Kim, HyeonseungIm [2012] identified computation whose skyline probabilities are higher than a given approach. They developed a probabilistic skyline algorithm called PSkyline which computes exact skyline probabilities of all objects in a given uncertain data set. PSkyline aims to identify piece of case with skyline probability zero, and more importantly, to find unmatched groups of case and assign with unnecessary power tests completely. To increase the chance of finding such pieces and groups of case, PSkyline uses a new in-memory tree structure called Z-tree. They also developed an online probabilistic skyline algorithm called O-PSkyline for uncertain data streams and a top-k probabilistic skyline algorithm called K-PSkyline to find top-k objects with the highest skyline probabilities.

Katja Hose, Akrivi Vlachou [2012] stated that skyline queries are necessary in order to help users to handle the very large amount of available data by identifying a set of interesting data item. Skyline query processing in highly distributed environments produce inherent challenges and demands and require current

techniques due to the distribution of content and the lack of total knowledge. This is interesting and still evolving research area so that monitor can easily obtain an overview of up to the minute. Outlined the objectives and the main rule that any distributed skyline approach has to fulfill, leading to useful guidance for developing algorithms for distributed skyline processing. They review in detail existing approaches that are suitable for highly distributed environments, clarify the as expectation of each approach and provide a qualified presentation analysis. Moreover, they study the skyline variation each approach supports. Their analysis leads to a classification of existing approaches. Finally, they present interesting research topics on distributed skyline computation that have not yet been discovered.

Yoonjae Park, Jun Ki Min [2013] has proposed efficient parallel algorithms for handling the skyline and its variation using Map-Reduce. They first frame bar graph to successfully lop out non-skyline (non-reverse skyline) points in advance. They next partition data based on the fields divided by the bar graph and compute candidate (reverse) skyline points for each area individually using Map-Reduce. Finally, they checked whether each candidate point is actually a (reverse) skyline point in every field individually. Their performance study checks the value and flexible of the proposed algorithms.

Ah Han [2014] a novel algorithm has been proposed to strongly process reverse skyline queries using an approach based on two pruning methods: the search-area pruning method and the candidate-objects pruning method. Using these pruning methods, the algorithm is intelligent to process reverse skyline queries easily even in situations where data is changing normally. The proposed algorithm also successfully reduces the invalid use of storage under existing approaches for storing estimate results. They presented general experiments to display that Pruning based reverse skyline (PBRS) algorithm shows better presentation compared to existing approaches compact of the measurement, distribution, or size of the data.

Zhiqiog Wang, Junchang Xin [2015] states uncertainty is the important characters identifying data, the processing and increase techniques for Probabilistic Skyline (PS) in wireless sensor networks (WSNs) are investigated. Distributed Processing of Probabilistic Skyline (DPPS) query in WSNs, is proposed. The algorithm divides the identifying data into candidate data (CD), irrelevant data (ID), and relevant data (RD). The ID in each sensor node can be filtered directly to reduce data transmissions cost, and then only allowing to both CD and RD, PS result can be properly found on the base station. Experimental results shows that their proposed algorithm can effectively reduce data transmissions by filtering the unnecessary data and greatly extend the lifetime of WSNs.

YunjunGao, Qing Liu [2016] causality and responsibility is a necessary tool in the database association for providing direct information for answers/ non-answers to queries. The authors analyzing the causality and responsibility problem (CRP) for the non-answers to probabilistic reverse skyline queries (PRSQ). They first explain CRP on PRSQ, and then, they proposed an valuable algorithm described as Causality and responsibility to Probabilistic reverse skyline queries (CP) to measure the causality and responsibility for the non-answers to PRSQ. CP first finds candidate elements, and then, it performs authentication to collect actual element with their responsibilities, during which several methods are used to advance qualification. Further, they analyze the CRP for the non-answers to reverse skyline queries. Almost they extend CP to identify directly all the certain elements and their responsibilities for a non-answer to reverse skyline queries without further verification. Major experiments using both real and fake data sets determine the performance and ability of their presented algorithms.

Following table shows advantages and limitations of skyline computing.

Table 1: Skyline Computing advantages and limitations

No	Author	Advantages	Limitations
1.	Dongwon Kim, Hyeonseung Im [2012]	• Minimum Description code length O-PSkyline is a lexicographic order provides better results.	• Extremely small number of patterns. • Not Scalable Failure detection takes lot of time overhead.
2.	Katja Hose, Vlachou A [2012]	• Works with highly distributed environment. • Better Embedding lists that provide good support for uncertain databases.	• Interesting patterns may be lost. • Extra overhead to check whether temporal patterns are closed.
3.	Yoonjae Park, Jun	• High Effective Pruning.	• Frequent graphs generated may not

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

	ki Min [2013]	• Good clustering with classification on high utility item mining. • Reverse computation check possible to verify accuracy of results.	be exactly frequent. • High computation time.
4.	AH HAN [2014]	• High effectiveness and scalable that reduce insufficient memory usage lexicographic order Better performances.	• Poor Time for summarizing the patterns is more than that for mining. • No better optimization in overall performance.
5.	Zhiqiong Wang, Junchang Xin [2015]	• Verified effectiveness on skyline query. • Data transmission cost is low. • It avoids irrelevant data which can affect query processing. • Similar value on taken to analyze with train and test data sets.	• Not suitable world explosion that function with multidimensional data. • Sensing time may increase due to uncertain data logs. • Large sensor needs more communication cost.
6.	Yunjun Gao,Qing Liu [2016]	• Support non answer queries with better results. • It can support both real and synthetic databases.	• Lack in time process to boost high dimensional data sets. • Indirect data centre queries result extraction is not possible.

III. CONCLUSION

The database studies community began to pay rising interest to the trouble of processing skyline queries. The recognition of the skyline operator is mainly due to its capacity to perceive a set of exciting items in a massive database. This paper focus on the P-skyline query in distributed environment, namely DSUD (Distributed Skyline queries over uncertain data) query. In order to accelerate the DSUD query, the Xu Zhou proposed an improved DSUD framework and designs an adaptive (ADSUD) algorithm. In ADSUD, several efficient technologies, including an IPR-tree and the reuse technology, are employed. Moreover, they define the MPBR (Minimum Probabilistic Bounding Rectangle) for collecting the global abstract information and selecting local representative tuples. Extensive experiments have been conducted to clarify the effectiveness and the efficiency of their algorithms. Considering MapReduce possess tremendous advantages in extracting, processing, and analysis of big datasets, the DSUD queries under MapReduce can be an extension work.

IV. REFERENCES

1 A. Wiley Brand, " Big data for Dummies" 2013.
2 K. Hose and A. Vlachou, "A survey of skyline processing in highly distributed environments," The VLDBJournal The International Journal on Very Large Data Bases, vol. 21, no. 3, pp. 359–384, 2012.
3 K. Dongwon, I. Hyeonseung, and P. Sungwoo, "Computing exact skyline probabilities for uncertain databases," Knowledge and Data Engineering, IEEE Transactionson, vol. 24, no. 12, pp. 2113–2126, 2012.
4 Y. Park, J. K Min, and K. Shim, "Parallel computation of skyline and reverse skyline queries using mapreduce," Proceedings of the VLDB Endowment, vol. 6, no. 14, pp. 2002–2013, 2013.
5 A Han, Youngbae Park and Dongseop Kwon+, "An Efficient Pruning Method to Process Reverse Skyline Queries," Journal of Information Science and Engineering vol.30, no 501-517(2014).
6 Zhiqiong Wang and Junchang Xin, "Alternative Tuples Based Probabilistic Skyline Query Processing in Wireless Sensor Networks," Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 813507.
7 Qing LIU and Yunjun GAO, "Finding causality and responsibility for probabilistic reverse skyline query non-answers," IEEE Transaction on Knowledge and Data Engineering, Vol.28, No.11, Nov 2016.
8 Xu Zhou and Kenli Li, "Adaptive Processing for Distributed Skyline Queries over Uncertain Data," IEEE Transactions On Knowledge And Data Engineering, Vol.*, No.*,* 2015.
9 Bin Jiang, Jian Pei, Xuemin Lin and Yidong Yuan, "Probabilistic skylines on uncertain data:model and bounding-pruning-refining methods" © Springer Science+Business Media, LLC 2010.

10 Maaruf Mohammed Lawal, Hamidah Ibrahim, FazlidaMohd Sani andRazaliYaakob "Skyline Computation of Uncertain Database: A Survey," Proceedings of the 6th International Conference on Computing and Informatics, ICOCI 2017 25-27April, 2017.
11 Samiddha Mukherjee and Ravi Shaw, "Big Data – Concepts, Applications, Challenges and Future Scope," International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016.
12 J. Pei, B. Jiang, X. Lin, and Y. Yuan, "Probabilistic skylines on uncertain data," in Proceedings of the 33rd international conference on Very large data bases. VLDB Endowment, pp. 15–26, 2007.
13 M. J.Atallah and Y. Qi, "Computing all skyline probabilities for uncertain data," in Proceedings of the twenty-eighth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. ACM, pp. 279–287, 2009.
14 W. Zhang, X. Lin, Y. Zhang, W. Wang, G. Zhu, and J. X. Yu, "Probabilistic skyline operator over sliding windows," Information Systems, vol. 38, no. 8, pp. 1212–1233, 2013.
15 J. B. Rocha-Junior, A. Vlachou, and C. Doulkeridis, "Agids: A grid-based strategy for distributed skyline query processing," in Data Management inGrid and Peer-to-Peer Systems. Springer, pp. 12–23, 2009.
16 Borzsonyi, Stephan and Konrad, " The Skyline Operator," Proceedings 17th International Conference on Data Engineering: 421-430,2001.
17 G. Trimponias, I. Bartolini, D. Papadias, and Y. Yang, "Skyline processing on distributed vertical decompositions," Knowledge and Data Engineering, IEEE Transactionson, vol. 25, no. 4, pp. 850–862, 2013.
18 Hailbert and martin. "Big data for Development: A Review of Promises and Chellenges. Development Policy Review.'' martinhilbert.net.Retrieved 2015-10-07.
19 Walunj Swaphil k,yadav Anil H and sonu Gupta, "Big data: Characteristics, challenges and data mining'' International Journal of Computer Applications(0975-8887). International Conference on Advance in Information Technology and management ICAIM-2016.
20 Melanie Luna "Big data management'' 2013.

CITE AN ARTICLE

Saravanapriya , S., & Thiagarasu, V., Dr. (2017). A SURVEY ON SKYLINE COMPUTING FOR UNCERTAIN DATABASE. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(9), 35-39. Retrieved September 4, 2017.