IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Data Updates on Streaming Data Warehouses Using XOR Based Key Expansion Algorithm

Helena Lawrence[*1], Anisha S S[2]
[*1,2] Dept.of Computer Science, TKM Institute of Technology, Kollam, India
helenalawrence96@gmail.com

Abstract

In streaming data warehouses advance updates are made when the new batch of data entered in to the warehouse. This is totally different from the traditional approach. Because the traditional approach data are update during down times. Here I am proposing the scheduling framework which will be handles all the problems encountered in traditional data warehouses. Proportional partition strategy algorithm is used for scheduling the data in data warehouses. This will overcome the EDF partition algorithm.XOR based key expansion algorithm is used to protect the data in the data warehouse. This algorithm is depending on the vigenere's matrix.

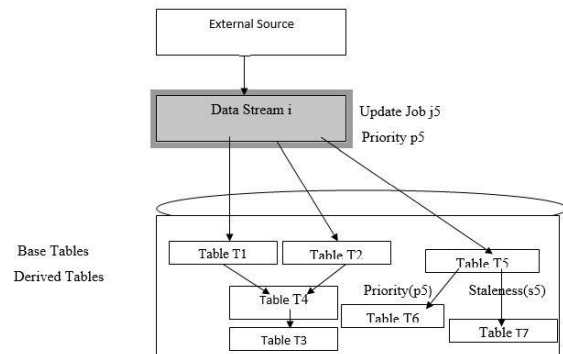**Keywords**: Scheduling framework, EDF algorithm. XOR based key expansion algorithm, VIGENER's matrix.

Introduction

Scalable scheduling of updates in streaming data warehouses which concludes the features of traditional data warehouses and streaming data warehouses. Data warehouse is nothing but storing large amount of data or large repositories to store data are called as data warehouses, there are two types of data warehouses. Traditional data warehouses and streaming data warehouses. The main aim of scalable scheduling is to propagate the data across all relevant tables and view very quickly as possible. one more aim is to reduce the data staleness over time. Multiple batch of data entered in to the data warehouse at the same time there is no mechanism to limit this simultaneous process, Here motivates the need of scheduler, so scalable scheduling is better choice to update the data in to the warehouse. Now a day's main problem in data warehouse is security. Information security is the challenges in data mining.Encrypt the data in such a way that data will be more secure.To encrypt the data we can follow XOR based key expansion algorithm.

Streaming Data Warehouse Architecture

This architecture mainly consist of two types of tables, base table and the derived tables, derived tables are the materialized views of the base tables. When the batch of data entered in the data warehouse first it will stored into the relevant base table. When some changes affected to the real time elements according to that we have to update this base table also. these changes will continue until all the derived

tables will change with the new data. after update the whole tables in the data warehouse, the updated tables will placed in to the queue, whose purpose is to execute the execute transform load mechanism.

Fig 1.1 Streaming data warehouse architecture

Scheduling Algorithms

Scheduling algorithms will work according to the scheduling frame work. It will meet all the constrains in the scheduling frameworks such as data consistency, transient overload, heterogeneity, non preemtability etc..Earliest deadline first algorithm(EDF) is the one of the dynamic scheduling algorithm, all the updates jobs are placed in to the priority queue, When some event will appear in the data warehouses such as finish update task releasing data ets..,. the queue will search for the job which is close to its deadline, This will be worked in hard real

time system so that the deadline should met with that particular time period.
Proportional partition strategy algorithm steps:
Create the cluster and add the jobs which have to update
Allocate the resources to each job
Now each job has each track
Jobs will execute in their tracks
It is a concurrent process and there is a separation from long job from the short jobs. A track logically represents the fraction of required resource to complete task.

Information Security

To achieve the security in data warehouse. Most challenges method is data encryption; it is the process of encoding the messages using a secret key. without know the key ,the hacker cannot hack the information. To generate the secret key I propose the XOR based key expansion algorithm. This is a symmetric encryption method. To encrypt and decrypt the data i am using the same key .User will provide one secret key. From this key I am generating the new key to encrypt the data. User provided key is not using for the data encryption. The full data can be divided in to the number of equal blokes and one key per block is applied.
Steps to generate the new key:-
Identify the user provided key and count the character length of the key.
Each character of the user provide key XORed and stored.
This result will be add with the length of the user provided key.
Again the result is XORed with the first block of the document.
This approach is differ from the traditional one because it is totally base on the ASCII values and vigenere's matrix. so the key as well as the encrypted message will be a alphabets only

Conclusion

The new approach using XOR based key expansion algorithm to encrypt the data includes alphabets, numerals. Special characters etc..New key is used for encryption as well as the decryption. This is the symmetric encryption method purely depend on the new generated key using XOR based expansion algorithm. Encrypt the data and stored in to the streaming data warehouses according to the proportional partition strategy algorithm.

References

[1] B. Adel berg, H. Garcia-Molina, and B. Kao, "Applying Update Streams in a Soft Real-Time Database System," Proc.ACM SIGMOD Int'l Conf. Management of Data, pp. 245-256,1995.

[2] Y. Huge, J. Wiener, and H. Garcia-Molina, "Multiple View Consistency for Data Warehousing," Proc. IEEE 13th Int'l Conf. Data Eng. (ICDE), pp. 289-300, 1986.

[3] J. Cho and H. Garcia-Molina, "Synchronizing a Database to Improve Freshness," Proc. ACMSIGMOD Int'l Conf. Management of Data,pp.117-128, 2000.

[4] L. Golab, T. Johnson, and V. Shkapenyuk, "Scheduling Updates in a Real-Time Stream Warehouse," Proc. IEEE 25th Int'l Conf. Data Eng. (ICDE), pp. 1207-1210, 2009.

[5] B.Babcock, S.Babu, M.Datar, and R.Motwani, "Chain: Operator Scheduling for Memory Minimization in Data Stream Systems," Proc.ACM SIGMOD Int'l Conf. Management of Data, pp. 253- 264, 2003.

[6] E. Anupriya, Sourabh Babelay, Amit Agnihotri and Sachin Soni "Encryption using XOR based Extended Key for Information Security – A Novel Approach", International Journal on Computer Science and Engineering (IJCSE) of ISSN: 0975-3397 Vol. 3 No. 1 Jan 2011