A Fast Approach to Improve the Security of Range Amalgamate Queries

V. K. Shanmuga Priya¹ & G. Mahalakshmi²
¹ME CSE NPR College of Engineering and Technology,
²Assistant professor & NPR College of Engineering and Technology.

Abstract- Admin can be used to login and registration process can held on the registration form. The user if the registered user can entered again OTP does not generated if the user entry is new means code can be send into his or her gmail account based on these way security can be improved. Accurate solution has to be provided in big data environment using hierarchical diffusion algorithm for partitioning the amalgamate queries. Datum which can be distributed into many form and then create a advanced assessment can be construct if any inquiry appear it achieve solution can be access exactly by encapsulate narrow estimate from all distribution using hierarchical diffusion algorithm. Conclusion can be originate detailed form. It run in heterogeneous environment. Range amalgamate queries produce result directly and reduced. Enormous amount of time consumption and produce result efficiently using this hierarchical diffusion algorithm and then k-means clustering algorithm can be used for clustering process. By using this k-means clustering algorithm to cluster the partitioned data into a estimate form. If any query arrives from user data can be provided by this estimate. In this way also used to improve fastness and provide a result efficiently in big data environment. AES algorithm can be used for improving security. Performance which can be evaluated and then generate the result. Result has to be encrypted and then uploaded and then using AES algorithm to decrypt the data and downloaded the data. Based on these way security and then fastness can be improved in big data environment.

Keywords: Advance Encryption Algorithm (AES), Datum, Hierarchical Diffusion Algorithm (HDA).

1. INTRODUCTION

Big data processing is the main task. In this processing some frameworks is Hive, pig, Jaql like technologies play an important role described. On the 6th kart announces an offer which is very cheap. Resulting in high sever processing is a very low small amount of time. According to Flip-kart there are billions of request hit within 30 min. For processing large amount of data and analyze that data various technologies are in use as mentioned above. Big data processing is the main task. In this processing some frameworks is Hive, pig, Jaql like technologies play an important role described. On the 6th kart announces an offer which is very cheap. Resulting in high sever processing is a very low small amount of time. According to Flip-kart there are billions of request hit within 30 min. For processing large amount of data and analyze that data various technologies are in use as mentioned above.

2. RELATED WORK

2.1 FastRAQ A Fast Approach to Range-Aggregate queries in Big Data Environments

Range amalgamate queries are apply a certain aggregate function on all tuples within given query ranges. Existing approaches to range aggregate queries are insufficient to produce results in big data environment. It divides data into indepent partitions with balance partitioning algorithm and generate local estimate sketch for each partition. The sampling and histogram approaches have been utilized in database environments to support approximate answering or selectivity estimation. However, it cannot acquire acceptable approximations of the underlying data sets, when data frequency distributions in different dimensions vary significantly.

2.2 Quantifying Trading Behavior in Financial Markets Using Google Trends

Crises in financial markets affect humans worldwide. Detailed market data on trading decisions reflect some of the complex human behavior that has led to these crises. They suggest that massive new data sources resulting from human interaction with the Internet may offer a new perspective on the behavior of market.
participants in periods of large market movements. By analyzing changes in Google query volumes for search terms related to finance, they find patterns that may be interpreted as “early warning signs” of stock market moves.

2.3 Predicting the Present with Google Trends

Search engine data to forecast near-term values of economic indicators. Examples include automobile sales, unemployment claims, travel destination planning, and consumer confidence. Government agencies periodically release indicators of the level of economic activity in various sectors.

3. IMPLEMENTATION

Hierarchical diffusion algorithm divides all data into different groups with regard to their attribute values of interest, and further separates each group into multiple partitions according to the current data distributions and the number of available servers. The algorithm can bound the sample errors in each partition, and can balance the number of records adaptively among servers when the data distribution and/or the number of servers changes. In each partition, the sample and the histogram are updated respectively by the attribute values of the incoming record. When a query request arrives, it is delivered into each partition. Query can be used to retrieve result from the big data environment. Security can be improved using AES algorithm using public key and private key for encryption and decryption process.

4. ARCHITECTURE

![Architecture of proposed system.](image-url)
5. MODULES

MODULE 5.1: Upload Dataset into Hadoop

In this module data which can be uploaded into the hadoop server. Fast Range Amalgamate Query has $O(1)$ time complexity for data updates and $o\left(\frac{N}{P \times B}\right)$ time complexity for ad-hoc range-amalgamate queries, where $N$ is the number of distinct tuples in all dimensions, $P$ is the number of partitions, and $B$ is the number of buckets in a histogram. Furthermore, it produces negligible volume of index data in big data environments. Hive is a typical data analysis tool with $O(N)$ time complexity for any ad-hoc range amalgamate queries. Hive can compile the task of an ad hoc range-amalgamate query into optimized mapreduce jobs and execute them on top of Hadoop.

MODULE 5.2: Split the Dataset

In this module data can be split by using hierarchical diffusion algorithm works with a stratified sampling model. It divides all data into different groups with regard to their attribute values of interest, and further separates each group into multiple partitions according to the current data distributions and the number of available servers. The algorithm can bound the sample errors in each partition, and can balance the number of records adaptively among servers when the data distribution and/or the number of servers changes. In each partition, the sample and the histogram are updated respectively by the attribute values of the incoming record. When a query request arrives, it is delivered into each partition. It first build cardinality estimator (CE) for the queried range from the histogram in each partition. To make data balanced on each server, the partition algorithm subdivides each group into a number of partitions according to the current data distributions and sends each partition onto one server.

MODULE 5.3: Build the range amalgamate queries

In this module splitted data can be clustered by using k-means clustering algorithm. The system use the common K-Means clustering method to analyze the vectors set and produce K clusters. A unique Cluster ID is assigned to each cluster. They construct a list of key-value pairs from the result of K clusters. The key-value pairs are in the format of $<\text{tag;}\text{Cluster ID}>$. They sort the key-value pairs by tag in alphabetical order. The buckets in the histogram are built from the sorted pairs. The key idea is to merge the pairs with the same Cluster IDs into the same bucket. If some tag occurring frequency is significantly different from others, its Class ID is different after the K-Means clustering, and it will be put into an independent bucket in the histogram.

MODULE 5.4: Execute the Range Amalgamate Query

In this module query can be executed. Fast Range Amalgamate Query uses approximate answering approaches, such as sampling, histogram, and cardinality estimation etc., to improve the performance of range-amalgamate queries. They use relative error as a statistical tool for accuracy analysis. Relative error is widely used in an approximate answering system. Also, it is easy to compute the relative errors of combined estimate variables in a distributed environment for Fast Range Amalgamate Query. In this section, to analyze the estimated relative error and the confidence interval of final range-amalgamate query result.

MODULE 5.5: Performance Analysis

In this module can evaluate the query cost and then time difference between the query process execution. First, this proposed work analyze the performance of Balance Hierarchical diffusion and Hierarchical diffusion algorithms. They examine the query cost of Fast Range Amalgamate Query. According to Algorithm, the records can be loaded to the servers with balanced load distribution. The queries operations can be carried out between partitions parallelly. Second, they analyze the update cost of Fast Range Amalgamate Queries. The updating process includes delivering a record to a specified partition, and updates the parameters of the histogram in a partition.

MODULE 5.6: Improving Security

Security can be improved using AES algorithm. Executed result can get as a data and it can be encrypted. Data can be encrypted using AES algorithm. Encrypted data can be send to another server and then data which can be decrypted based on this process security can be improved on query processing in big data environment.
6. CONCLUSION

The goal of the system is to improve the security by using advance encryption algorithm which contains two types of keys private and public key which can be used for encryption and decryption process. Result can be produced fastly using hierarchical diffusion algorithm in big data environment query can be used to retrieved result based on this process fastness can be improved in big data environment. Fastness with security process can be improved in big data environment.

REFERENCES


