# A Survey on Batch Stream Processing using Query Replacement Technique to Avoid Database Risks

**Pranav Talwar[1], D. H. Kulkarni[2]**

[1]Pune University, Smt.Kashibai Navale College of Engineering, Vadgaon(BK), Pune411041, India

[2]Pune University, Smt.Kashibai Navale College of Engineering, Vadgaon(BK), Pune411041, India

**Abstract:** *Due to drastic increasing in web applications and also internet users, there is huge number of transactions are happening at server side. This in turn affects the database too. So when database is crowded with huge number of queries then database server becomes so busy to handle these queries. So either these query will be executed in some delay of time or they vanish with the object over weight. So proposed method is introduces an idea of using fuzzy logic with the batch stream processing of the queries which are stored in the priority queue for avoiding database performance degradation.*

**Keywords:** Fuzzy logic, Batch stream processing, Query replacement, Database Management

## 1. Introduction

Due to globalization of industries over wide range of area, it has become essential for enterprises to extend their work-force in order to meet the requirement. As a result the volume and complexity of enterprise data also grow. The information that are generated, processed, transfer, and store by an enterprise plays a vital role among the asset of an enterprise. In today's data-driven world, Organizations often use IT systems to store and retrieve vast amount of sensitive data, which, if leaked, could be harmful to an organization and may lose its competitive position in the market.

However, it is seen that system performance is highly dependent on different queries and the data within the system that constitute overall load rather than system size. Many organization uses cloud for storing, retrieving and manipulating data, because of its ease of access, integration, portability, and security issue in corporate world. Issue that lies within many enterprises is to deal with the ability to meet their rising demand of work while reducing their operational costs of hardware. To overcome the drawback we have to define effective measures that guarantee a wise usage of computing resources.

Information play a vital role in any organization and management of data and preserving privacy is main concern of organization. Database management not only provide information retrieval but also deploy data integrity of the stored data.

## 2. Literature Review

To ensure secure data integrity [1] presents misuse detection paradigm called DEMIDS which is made-to-order to relational database systems. DEMIDS uses audit log to keep track on profile of user interaction with DBS. By keeping track on profile we can detect misuse behavior, in particular insider abuse. Although DEMIDS is helpful to detect both interruption and insider abuse, DEMIDS play an important role in detecting the mischievous behavior by legitimate users who break their privileges. Hence DEMIDS is particularly useful for internal control. DEMIDS also

considers the data structure and semantics species present in the database schema through the view of distance measure. Such system data is used to guide the frequent item sets to search for frequent item sets which can effectively represent the working scopes of users.

To enable access control over database system Role Base Access Control (RBAC) facilitate the user to handle the problem of illegal access to unauthenticated user.[2] abbreviate a model of RBAC that introduces a novel medium to measure the risk incurred with role misuse. It extends the role of RBAC that provide valuable security for access control. Depending upon each role we assign different risk level that determine which user take what action on existing database.

Communication and Collaboration is an important aspect of information retrieval (IR) system. This concern is address by sharing other user search experience which help to formulate user queries .Approach [3] introduce, a collaborative querying paradigm that helps users with query formulating by finding previously submitted similar queries through mining. Here in collaborative query approach a group of similar queries cluster is formed atomically without predetermined class description. Experimentally a hybrid content-based and results-based approach, both of which give better query clusters result , than using either of them alone .

Most of databases administrators reduce workloads by aborting or pausing a queries and thereby increase system performance. such a view is not efficient for several reason like, aborting queries that may use unwanted resources without satisfying the need of user queries. Our decision is accomplished by evaluating these view of formulating queries and taking appropriate actions we can minimize impacts of resource demanding queries on some systems. Do deal with this, we must know the level of acceptance of the systems .[4] adapted a novel base approach which incorporate Multi–Level Security (MLS) access control model that is used to develop a risk-adaptive access control model. This new model is known as Fuzzy Logic control

system which can even allow risky information needed by a user, provided the risk can be controlled.

[5] Exploits, available information present in the logs in order to measure the semantic similarity between the queries from user and that present in the logs. It focuses on the concept of query flow graph which is used for graph-based representation of a query log. The query flow graph integrates query reformulation for many clients, similarity between query is obtained by extracting appropriate sub graph on a low Euclidean space .Further, the resultant query similarity is use for application like query recommendation. In order to share information over a dynamic and distributed network, an attempt has been evolved for risk base access control system. To overcome the above problem of risk through the network,[6] developed a framework based on Risk Base Access control mechanism that take decision on risk associated with operational need and given request. The propose framework estimate risk and impact score which depend on object sensitivity-based approach and a subject trustworthiness-based approach, further two additional approaches that are used for determining threat ordering are based on the difference between object sensitivity and subject trustworthiness.

[7] develops a model that defines a problem of insider attack. The insider attack address toward the risk assessment policy of any organization that are bounded with security policy, that indicate which user are authorized to access what resources and when. Their work contributes to analyze the statics behavior of user capabilities and restriction of given policy at runtime. Often times user do not get a satisfactory result over the fired query which may result in unsatisfactory result.[8] demonstrate a framework that depict the query over database in domain independent way .They present AOIM a user and domain dependent approach for answering a relevant queries from database. It efficiently extract and automatically rank the tuples that satisfy query and the structure of the relations projected by the databases. A test has been carried out in term of satisfying user by considering only the information contained in the dataset. AOIM has been implemented without affecting the internal structure of database and therefore it can be easily implemented over any web database.[9] narrates a generic automated ranking paradigm for SQL database which implement a ranking algorithm for relational DBMS. It compromise of two major component query processing component and pre- processing component that minimize the impact on query processing.

As a enterprise grow the functional requirement and administrative domains also spans with time. Control policies o an enterprise typically express the requirements of an authorization constraints.[10] presents a work on X-GTRBAC Admin that focuses on administration model that aims at enabling the role of administration in role-based access control (RBAC) policies in the presence of constraints. These model represents a solution to the administration problem for enterprise-wide access control, which not only includes access control management for users but also allow decentralization of administration policy tasks through the abstraction of administrative domains resources within a single domain. [11] describe , a reference base model for RBCA, over past decade it has seen that many

system result in confusion of its utility and meaning. To overcome the drawback NIST model resolve the situation by taking idea from prior RBAC model .Here the system is organize into four level of RBAC namely flat ,hierarchical, constrained and symmetric RBCA with increasing functional capabilities.[12] depicted a efficient model for predicting the essential query over the database. It proposed a prediction method for unstructured data to measure the difficulty of query over database using ranking principle. This method predict the difficulty of query with low errors overheads and structured robustness score based on the query similarity between the original query and the corrupted version of same query present in the database. It uses both the INEX and Research benchmarks for keyword search on database that predict the difficulty of query efficiently and effectively.

[13] convey a method for the automated segmentation of users' query streams into hierarchical units. This system uses a set of query log, setoff syntactic ,temporal and web search features that can predict goal and mission boundaries as well when use independently .This approach adopts a hierarchical model for grouping a set of queries that will allow user more precisely found the required data over a paradigm. It allow us to more accurately measure time required for user queries to complete its task. User query is not evaluated on a per-query basis, but on the basis of user tasks. A cooperative approach was developed for fuzzy relational queries to avoid empty answer to the user.[14] proposed a relaxation mechanism for more complex queries .This mechanism assist On transformation by applying a query tolerance to fuzzy predicate contained in a query ,which can be conveniently used in terms of parametrized proxy relation. Predicate of fuzzy is obtained by simply performing arithmetic operation on fuzzy numbers. It is concluded that a proximity relation is define in a relative notion which play a key role.

## 3. Conclusion

The proposed approach by using fuzzy logic successfully identifies the database server performance parameter and it break down points. Based on these things our system identifies the high priority queries from the priority queue depended on the extracted features. System introduces an idea of clustering these queries from the priority queue based on the feature score. Then these queries are batched based on the query type to commit in single go to yield best performance of the database even in the on the high risk scenario.

## References

[1] K. Levitt C. Y. Chung, "M. Gertz. Demids: A misuse detection system for database systems", In Proceeding of the Integrity and Internal Control in Information System, 1999.
[2] EbruCelikel, Murat Kantarcioglu, BhavaniThuraisingham, DilsadCavus," A Model for Risk Adaptive Access Control in RBAC Employed Distributed Environments",University of Texas at Dallas Richardson, TX 75083,2007
[3] Lin Fu, Dion Hoe-Lian Goh, Schubert Shou-Boon Foo, YohanSupangat," Collaborative querying for enhanced

Paper ID: SUB1582

information retrieval",Research and Advanced Technology for Digital LibrariesLecture Notes in Computer Science Volume 3232,2004

[4] Pau–Chen Cheng, Pankaj Rohatgi Claudia Keser ,Paul A. Karger, Grant M. Wagner Angela SchuettReninger,"Fuzzy Multi–Level Security : An Experiment on Quantified Risk–Adaptive Access Control"Security and Privacy, SP '07. IEEE Symposium on 2007.

[5] C. Castillo A. Gionis I. Bordino, D. Donato. "Query similarity by projecting the query-flow graph". SIGIR'10, July 19-23 2010.

[6] VenkataBhamidipati ,SavithKandala, Ravi Sandhu. "A framework for threat assessment in access control systems" that appeared in Proceedings of 27th IFIP TC 11 Information Security and Privacy Conference (SEC 2012), 2012.

[7] Jason Crampton, Michael Huth," Detecting and Countering Insider Threats: Can Policy-Based Access Control Help", Proceeding of 5th International Workshop on Security and Trust management,2009

[8] S. Kambhampti U. Nambiar," Answering imprecise database queries: A novel approach", WIDM, New Orleans, Louisiana, USA, 2003.

[9] G. Das A. Gionis S. Agrawal, S. Chaudhuri, "Automated ranking of database query results" Proceedings of the 2003 CIDR Conference.

[10] E. Bertino A. Ghafoor J. Joshi R. Bhatti, B. Sha_q. Xgtrbac admin" A decentralized administration model for enterprise-wide access control." ACM Transactions on Information Security 8, pages 388{423,2005}.

[11] F. D. Ferraiolo R. Sandhu and D. R. Kuhn. "The nist model for role-based access control", Towards a unied standard. Information Technology Lab, NIST, September 2000.

[12] Shiwen Cheng, ArashTermehchy, and VagelisHristidis," Efficient Prediction of Difficult Keyword Queries over Databases", Knowledge (Volume:26 , Issue: 6 ),2013.

[13] Rosie Jones ,Kristina Lisa Klinkner,"Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs", Proceeding CIKM '08 Proceeding of 17th ACM conference on Information and knowledge management,2008.

[14] patricbosc,allelhadjali and olivierpivert, "Towards a tolerance base technique for cooperative answering of fuzzy queries against regular databases", springer verlag berlin heidelberg,2005.