Query Optimization of Employee Performance in Supply Chain Management System

Eswaran Badrinarayanan Venkateswaran
Hani Pankajkumar Bhavsar
Vishal Jayaraman

COMP 8157: Advanced Database Topics
Project Report

Submitted to: Dr. Kalyani Selvarajah
Term: Winter 2022
School of Computer Science
University of Windsor

Query Optimization of Employee Performance in Supply Chain Management System

Eswaran Badrinarayanan Venkateswaran[1], Hani Pankajkumar Bhavsar[2], Vishal Jayaraman[3]

[1, 2, 3] Graduate Student, School of Computer Science, University of Windsor, Windsor, ON N9B 2E7

[1]badrina@uwindsor.ca, [2]bhavsa92@uwindsor.ca, [3]jayaram3@uwindsor.ca

I. ABSTRACT

Advancements in technology have facilitated established companies to significantly manage stocks, inventories. Modern-day computer machines have been programmed to automate concurrent processes, keep track of the goods and services, operate seamlessly to produce large-scale sales and impact the business growth of organizations. In addition, they have contributed to shaping up novel approaches and strategies in terms of reforming the structural process of Supply Chain Management (SCM). However, businesses that run at a small scale need an effective and efficient method to develop their market base, analyze business trends, estimate financial budgets, collaborate with clients and attract potential customers. Managers and Supervisors need an effective system to search and fetch historical employee records to track employee performance. The project helps to overcome potential constraints associated in establishing the business obtained from the numerical metrics and supports the concept of entrepreneurship by harnessing costs, miscellaneous expenditure to resource efficiency aided by an efficient management and analysis system to effectively impact and enhance Canada's growth in the economy drastically affected by Covid-19 that would preferably specify a competitive platform for emerging business start-ups over the next few years and motivating youths to consider entrepreneurship as a prospective career.

Keywords: **Query Optimization, Indexing, Tableau, MS SQL Server, Analysis, Employee Performance.**

II. INTRODUCTION & MOTIVATION

As Entrepreneurship is seen as one of the driving factors behind creativity, profitability, employment opportunities, and economic progress in Canada, Countries with higher level of entrepreneurial activity are often economically prosperous [1]. Investors have had a tremendous impact on Canadian civilization, and today, an increasing number of Canadians from across all kinds of backgrounds are becoming or are contemplating becoming entrepreneurs. Canadian entrepreneurs are well-liked in their neighborhoods and in the media. The project work addresses the problem of time consumed to fetch historical employee records. At the same time, the league of entrepreneurship is challenging. Emerging businesses must deal with volatile markets, untested technologies, and inconsistencies in organizational processes. Furthermore, information insights about the management's production will help to better predict futuristic goals and set grounds for its accomplishment. Market competitions between businesses can be examined and compared against different metrics. Thus, it encourages small-scale start-ups to step forward and provides approaches to stay on par with leading commercial organizations. Thus, this will initiate motivation in the minds of budding young entrepreneurs wherein they can be benefitted by deducing analysis for the successful growth of their businesses.

The fate of entrepreneurial enterprises is primarily decided by business environment features such as availability to finance, access to international partnerships, and consumers' purchase power, over which individual entrepreneurs have limited control. As a result, the global survival rate of small enterprises is low, and the growth rate is much lower. Considering the importance of entrepreneurship in Canada, it is essential to determine how successfully Canadian entrepreneurs are addressing these difficulties.

The purpose of query optimization is to decrease the system resources necessary to satisfy a query, allowing the user to receive the proper result set faster. For starters, it gives the user faster results, making the program appear faster to the user. Second, because each request takes less time than unoptimized queries, it allows the system to handle more inquiries in the same amount of time. Finally, query optimization decreases the amount of wear on hardware (e.g., disc drives) and allows the server to run more efficiently (e.g., lower power consumption, less memory usage).

Considering one of the business sectors, the Food Industry which in terms of the value of production and sales of manufactured products is the second biggest manufacturing industry in Canada, wherein automation software services such as Clearview are employed in Quick Service Restaurants (QSR) that has largely influenced managerial operations. However, the existing system lacks the transparency of information release to employees, access to data backups (storage of data) and a predictive analysis system for financial budgeting based on resource usage.

According to SME reports (Statistics Canada), the following areas could be reinforced moving ahead in time [2];

- Canadian businesses that are fast-growing are generated at a lower proportion in the service sectors when compared globally with other countries.

- The owners of business organizations are aging in population and over the next few years need to consider succession planning and ownership transfer to the upcoming generation.

The report also identifies key points of consideration in the following aspects of activity concerning entrepreneurship that require more data to be studied further and they are as follows;

- Existing data do not allow for an examination of the breadth and diversity of entrepreneurship in Canadian firms and communities.

- Critical Indicators such as age, employment rate is often absent in entrepreneurship activity and performance analysis.

Besides, the firms having less than 250 employees possess limited practical knowledge related to Canadian entrepreneurship.

The project system involves primary database concepts like uploading and retrieval of data, fetching individual records, generating performance reports through data normalization and query optimization that are closely related to Advanced Database Topics.

III. BACKGROUND STUDY & RELATED WORKS

In the paper Query Optimization in Database Systems, the authors Jarke, M. & Koch, J. discuss about logic-based and semantic transformation approaches of query optimization in database systems [3]. Given a query, there are many access plans that a database management system (DBMS) can follow to process it and produce its answer. All plans are equivalent in terms of their final output but vary in their cost, that is, the amount of time that they need to run. This cost difference can be several orders of magnitude large. Thus, all DBMSs have a module that examines “all” alternatives and chooses the plan that needs the least amount of time. This module is called the query optimizer.

Efficient methods of processing unanticipated queries are a crucial prerequisite for the success of generalized database management systems. A wide variety of approaches to improve the performance of query evaluation algorithms have been proposed: logic-based and semantic transformations, fast implementations of basic operations, and combinatorial or heuristic algorithms for generating alternative access plans and choosing among them. These methods are presented in the framework of a general query evaluation procedure using the relational calculus representation of queries.

In addition, nonstandard query optimization issues such as higher-level query evaluation, query optimization in distributed databases, and use of database machines are addressed. The focus, however, is on query optimization in centralized database systems.

Exact optimization of query evaluation procedures is in general computationally intractable and is hampered further by the lack of precise statistical information about the database.

Query evaluation algorithms must rely heavily on heuristics. Nevertheless, the term "query optimization" will be used to refer to strategies intended to improve the efficiency of query evaluation procedures and reduce the search effort required to derive a particular record.

The economic principle requires that optimization procedures either attempt to maximize the output for a given number of resources or to minimize the resource usage for a given output. Query optimization tries to minimize the response time for a given query language and mix of query types in a given system environment. This general goal allows a number of different operational objective functions.

The response time goal is reasonable only under the assumption that user time is the most important bottleneck resource. Otherwise, direct cost minimization of technical resource usage can be attempted.

Fortunately, both objectives are largely complementary; when goal conflicts arise, they are typically resolved by assigning limits to the availability of technical resources (e.g., those of main memory buffer space).

Chen, Z., Gehrke, J., & Korn, F in their work on Query Optimization in Compressed Database Systems present an effective approach for compression based on lightweight, attribute-level compression techniques [4]. The authors demonstrate the Hierarchical Dictionary Encoding strategy that intelligently selects the most effective compression method for string-valued attributes. The authors also show that eager and lazy decompression strategies produce suboptimal plans for queries involving compressed string attributes.

The authors formalize the problem of compression-aware query optimization and propose one provably optimal and two fast heuristic algorithms for selecting a query plan for relational schemas with compressed attributes. In addition, the authors state that their results highlight the order speed up in CPU time over existing approaches.

McHugh, J., Widom, J., Abiteboul, S., Luo, Q., & Rajaraman, A. in the paper on Indexing Semi Structured Data techniques [5] describe techniques for building and exploiting indexes on semi structured data: data that may not have a fixed schema and that may not be regular or incomplete. Initially, the authors represent the solution as a general framework for indexing values in the presence of automatic type coercion and later introduce four indexes used for query processing based on a DBMS for semi structured data (Lore).

The authors create Lore from the scratch which allows to investigate how semi structured data affects each component of a database management system. As the paper focuses more on indexing, the tradeoff between efficient query performance and space and update cost must be considered in each DBMS. By basically recreating elements of the database in special-purpose structures, indexing enables rapid access to data.

While on the other hand, Roussopoulos, N. in his research on View Indexing in Relational Databases [6] resolves the problem of selecting an optimal index for efficient retrieval of views.

Although, the previous related works provides an optimal solution at most cases, a further analytical and insightful study is not often accompanied with query optimization. Henceforth, our project combines the queried results fetched through SQL query tuning techniques, indexing and are fed into the BI tool (Tableau) for analysis.

IV. PROPOSED MODEL

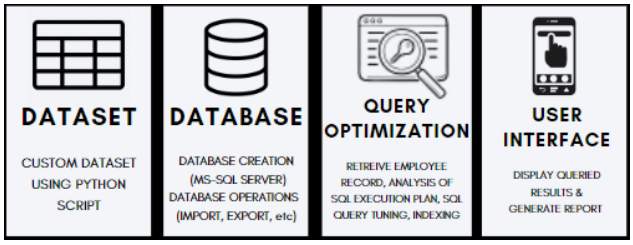


Figure 1. System Architecture

The proposed system is implemented as 4 step approach following a waterfall model of developmental cycle.

A. CREATION OF CUSTOM DATASET:

Python libraries such as faker and pandas are employed in this project. Faker is an open-source python package that allows users to create dummy datasets, or random data containing random attributes such as names, ages, localities, and so on. Pandas is a popular open-source Python package used for data science, data analysis, and machine learning. Pandas simplify several time-consuming, repetitive data-processing tasks, such as data normalization, merges and joins, data visualization, statistical analysis, data inspection, and data loading and storing.

In this project package is primarily used to generate dataframes and export them into spreadsheet format, which can then be imported into a database. With the help of the above-mentioned libraries a python script was written to generate huge number of synthetic data.

B. DATABSE DESIGN:

Microsoft SQL Server is used in this project as it deals with structured data. The main purpose of the database is to fetch the data of employee that are required to calculate the efficiency of the employee over a given period of time. Because the data is time series data the volume of the data is huge.

Therefore, the design of the database should be in such a way that;
1. The database should be normalized to the maximum level.
2. Implementation of query optimization is possible. By taking the above-mentioned conditions into consideration the database was designed.

There are three tables with the respective attributes in the database.
1. Employee - EID(Primary Key), Name, Email.
2. Product - PID(Primary Key), Name, Description.
3. Production - Date_of_Production, EID(Foreign Key), PID(Foreign Key), Produced_Quantity, Total_Quantity, Work_hours(Date_of_Production, EID(Foreign Key), PID(Foreign Key)(Primary Key). Here, a composite primary key is used as the date repeats in cycles of 10 for each employee record.

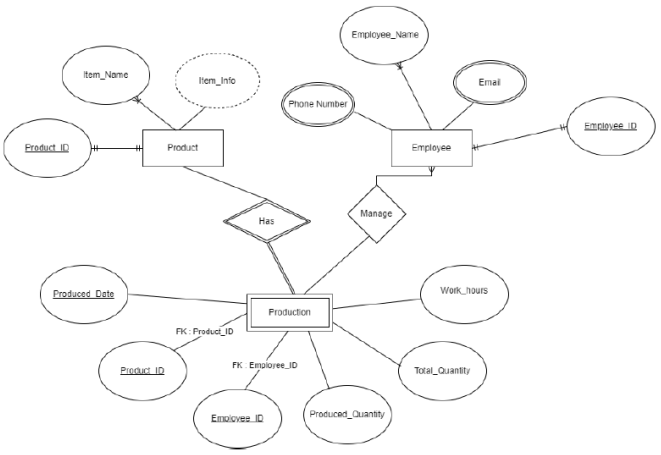


Figure 2. Entity Relationship Diagram

C. QUERY OPTIMIZATION:

The act of examining SQL queries and selecting the most efficient execution mechanism is known as SQL query optimization which is the significant advance database concept involved in retrieving employee record. It is usually a trial-and-error procedure in which many queries are examined to discover which one provides the greatest performance while still delivering the desired data [7].

Although query optimizers are occasionally included into database management systems, external, third-party solutions are frequently regarded to provide better performance results [8]. For each query, a typical query optimizer will create one or more query plans, each of which provides a method for running the query. Each query plan's performance (i.e., execution time) is measured, and the most efficient query plan is chosen and utilized to execute the query [9].

The team's key contributions towards the solution's outcome will potentially employ advanced database concepts such as query optimization, normalization on large scale employee records to derive individual performance efficiency based on query given parameters where the queried results will be generated as a report via email using a frontend UI that can be viewed by managers for further evaluation. The SQL query execution plans can be viewed and analyzed using an in-built feature of initial SQL execution plan and statistics. Originally, the team will effectively develop a custom dataset with curated related set of tables with python script to produce large-scale data.

The database fields will be created in the database server (MS SQL Server). The queried results will be generated in the shell terminal as a performance report to managers with a frontend user interface. Sessions (event handlers) are created to monitor the computation time, number of logical reads, etc. and aggregate functions are applied appropriately to estimate execution time between different queries. SQL Tuning techniques are also examined by deriving comparisons between like and equals. Index (non-clustered index) is created on production_date. The query cost and row scans are substantially reduced as a result of index creation.

D. DATA ANALYSIS

For analysis, the queried results are sent into BI tool. Tableau is used in the proposed model as it offers a user-friendly interface and supports interactive visualizations with multiple scripting languages. Also, Tableau can handle millions of rows of data with ease. Different types of visualization can be created with a large amount of data without impacting the performance of the dashboards and users can make live to connections to different data sources like SQL Server etc. The views created in SQL Server namely Efficiency and Employee Efficiency along with table data from Employee, Product and Production tables are imported into Tableau. Inner Join and relationships are established with common key fields to analyze data. The fields of different tables and views are labelled as Dimensions and Measures respectively. Dimensions are discrete values which does not change over time and remain constant throughout for e.g.; name, email, etc. On the other hand, Measures are continuous values (quantitative variables) that can be subjected to changes accordingly with time for e.g.; work hours, produced quantity. Heat Maps are generated with values from different tables and rows, columns are manipulated accordingly.

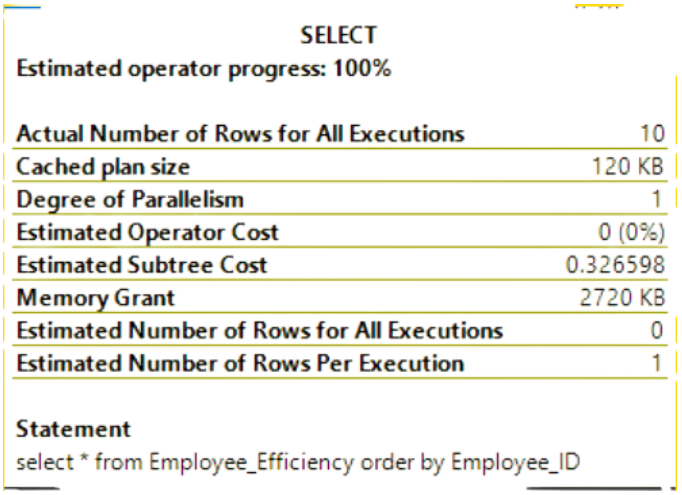
The calculated field is applied on field avg successful produced quantity using if-else condition statements with three feedbacks namely below average, on par average and above average for each of the employee's average product produced per hour. Explain data feature is also used to better understand behavioral patterns of employee data. Stories are created for different heat maps that are generated for interpretation of analytical data. The actions feature in tableau is incorporated to generate tailored email reports to each employee.

V. RESULTS

The subtree cost represents the estimated cost of a plan. It can be useful when investigating why the query optimizer chose one plan over another [10]. In our project, the subtree cost before indexing shows 0.33 (approx..) for a view select query involving multiple inner joins as displayed in Figure 3. While after indexing is performed on the same query, the subtree costs reduce to 0.12 (approx.) as indicated in the figure 4. The subtree cost is the overall cost required to run a query. The figure displays the subtree cost of the query executed that illustrates how the notion of indexing in the query optimization is efficient in reducing the overall subtree cost. SQL Query Tuner uses database profiling to rapidly and simply identify SQL queries that cause poor database performance and displays a graphical representation of wait time analysis [11].

It will continuously profile an entire data source over a specified time period. SQL Server Database Engine Tuning Advisor (DTA) can be used for debugging a specific issue query's performance, adjust a large group of queries across one or more databases, do an exploratory what-if examination of prospective physical design modifications, and handle storage space [12].

Database tuning is often referred to as performance tuning, and it entails the optimization and the homogenization of the database's design files and environment [13]. This improves the accessibility of data and helps managers; supervisors can fetch faster results of employee records as an outcome.



.

Figure 3. Subtree Cost before Indexing for a View Query

The programming language that the project will involve is Python. Since the project includes analysis as one of the key components, it provides support to data visualization and statistical computation of data. Also, it has fewer syntactical keywords, increased code readability, faster processing speed and facilitates automation of tasks. The database management choice will be relational DB - MS SQL Server as the concept of query optimization can be better demonstrated with set of relational tables and structured data. In addition, ODBC can be made between SQL server and Python using Pyodbc that can further be explored for an insightful analytical study.

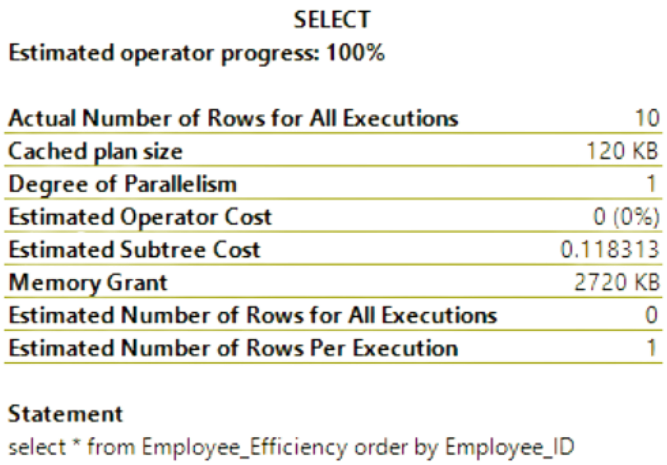


Figure 4. Subtree Cost After Indexing for a View Query

Figure 5 displays the heatmap analysis which is the process of evaluating and analyzing heat map data to acquire insights about employee engagement and behavior as they engage with the product done using the tableau software. Figure 4 displays the email that is automatically created by the tableau using the actions feature.

Figure 5. Heat Map Analysis using Tableau

In the above figure, the employee located in Alberta with an avg product produced per hour of 12.34 and avg successful produced quantity of 73 percent approx. for the item soups.

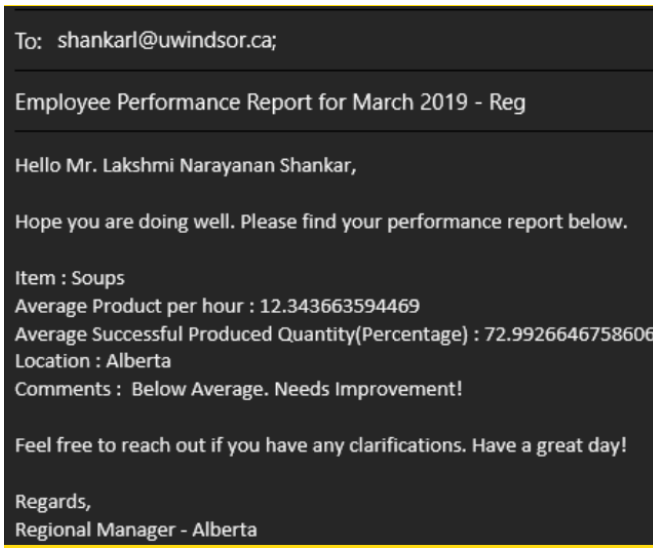


Figure 6. Generating Email Report

The email report is customized for each employee automatically by taking values from the tables and views. The actions feature uses regex patterns to identify and receive dynamic values for each field. The value delimiter and line delimiters are specified to add multiple field values. This can potentially serve as monthly report (performance metric) for all employee in a small-scale business. Personalized comments for individual employees are given with the help of calculated fields. Personalized comments can transform the employee performance of small-scale businesses as they are in the growing phase, each initiative taken can contribute to the final outcome.

VI. CHALLENGES/LIMITATIONS

There were several challenges encountered throughout the project's implementation. Organizations, governments, clicks, online servers, business partners, and even the human body generate data. However, no data was accessible online since the project was one-of-a-kind and only served as a simple demonstration of query optimization in SCM databases. As a result, dummy data was developed, which is not as realistic as a real-time dataset. Even while the waterfall approach is the greatest fit for this project, it does present some issues.

The waterfall model is fully based on a step-by-step approach to project management. This provides very little flexibility for future modifications. It is very difficult to execute an unanticipated modification in the future. It is risky to delay the testing phase till the end of a project, yet Waterfall forces teams to wait until the last step to test their products. The code will be tested later, when the queries have been optimized. Because the project has most likely taken the assigned amount of time to finish, major adjustments might create severe delays.

If the custom data created for the project does not cover all of the edge circumstances. This may lead to the project having some bugs. It is critical that the data be normalized throughout the database design and development process. ACID attributes are used to ensure that the database adheres to the requirements set forth by the database designer. A substantial challenge arises because the database will not adhere to the ACID characteristics if the normalization procedure is not completed correctly.

Given that one of the primary goals of the project is to develop query optimization, the amount of data created, as well as the number of participants, is large. A substantial likelihood exists that poor indexing will be carried out, which may have a negative impact on performance. Consequently, queries that are inefficient may be generated as a consequence of this. Aside from that, the project was built according to the waterfall model, which made it considerably more difficult to make adjustments later on. The optimization team and the testing team were forced to wait until the data generation and database design teams finished their work before they could proceed.

VII. CONCLUSIONS & FUTURE WORK

The project's aim of demonstrating the influence of query optimization on employee performance in the SCM DB was successfully accomplished via the analysis of the data collected. The main unique selling points of this project is the use of a tableau worksheet to visualise and analyse the dataset, as well as the automated emailing of the employee by the manager. Heat maps were used to illustrate the data, and the employee got email feedback that was automatically created and matched to his or her needs. Indexing is a considerably more efficient technique for searching through a large dataset than searching through a small dataset.

It helps small businesses save money by lowering their subtree costs. It also helps them save money by lowering the expense of searching through their information.

In spite of the fact that this project provided the fundamentals for a system, the first significant improvement would be the application of transformation rules (for example, relational algebra), which may be used to optimise queries in order to decrease the cost of processing queries. In addition, multi-level indexing on huge datasets may be utilized to get results more quickly and efficiently. It is possible to build Query Stores and Event Sessions in order to track query statistics and performance.

When complex industrial information is obtained in real time, supply chain analysis may be done, and the findings can be utilized to give precise and pragmatic answers in a range of circumstances. In order to accommodate semi-structured and unstructured data (which is more prevalent in the contemporary age) for query processing, database design may be made flexible. This is more typical in the contemporary age.

For more important insights into the dataset under examination, it is feasible to dive further into advanced tableau features like data blending, predictive analysis, and other similar techniques. A prototype for a more visually appealing and intuitive template for the matrix manager was developed and then implemented in the system. This new, more user-friendly design has the potential to be used across the whole system. Finally, it is necessary to have a broad query capability. When employees are asked to fill out product surveys, when managers or supervisors ask them questions, or when they need to gather information for different reports, this feature will be very useful for all of these things.

In conclusion, the project's goal to demonstrate the impact of query optimization on employee performance in the SCM DB was successfully analyzed. Heat maps were utilized to visualize the data, and the employee received autogenerated tailored feedback in the form of an email.

In future, Query tuning can be used to optimize queries which will improve the performance of the DB using transformation rules (Relation Algebra, eg: Nested Queries), multi-level indexing on large datasets can be used to fetch faster results, Query Store and Event Sessions can be created to monitor query statistics and performance. Supply chain analysis can be conducted on complex industrial datasets collected in real time that may be utilized to provide precise pragmatic outcomes.

Database design may be made flexible to handle semi-structured and unstructured data (which is more common in the modern era) for query processing. Advanced tableau features, such as data blending, predictive analysis, and so on, may be explored in more detail to provide better meaningful insights on the dataset.

VIII. REFERENCES

[1] Small Business Tourism and Marketplace Services. (2021, December 15). Key Small Business Statistics - SME research and statistics. Statistics Canada. https://www.ic.gc.ca/eic/site/061.nsf/eng/h_02689.html

[2] Overview of the food and beverage processing industry - agriculture.canada.ca. (2021, October 28). Agriculture Canada. https://agriculture.canada.ca/en/canadas-agriculture-sectors/food-processing-industry/overview-food-and-beverage-processing-industry

[3] Jarke, M., & Koch, J. (1984). Query optimization in database systems. ACM Computing surveys (CsUR), 16(2), 111-152.

[4] Chen, Z., Gehrke, J., & Korn, F. (2001, May). Query optimization in compressed database systems. In Proceedings of the 2001 ACM SIGMOD international conference on Management of data (pp. 271-282).

[5] McHugh, J., Widom, J., Abiteboul, S., Luo, Q., & Rajaraman, A. (1998). Indexing semistructured data. Stanford.

[6] Roussopoulos, N. (1982). View indexing in relational databases. ACM Transactions on Database Systems (TODS), 7(2), 258-290.

[7] Yevtushenko, A. (2019, April 24). Supercharge Your SQL Queries for Production Databases. Sisense. https://www.sisense.com/blog/8-ways-fine-tune-sql-queries-production-databases

[8] Thalheim, Bernhard. (1993). Database Design Strategies. 267-285. 10.1007/978-3-7091-2704-9_14.

[9] Khan, M., & Khan, M. N. A. (2013). Exploring query optimization techniques in relational databases. International Journal of Database Theory and Application, 6(3), 11-20.

[10] Zheng, J. G. (2017). Data visualization for business intelligence. Global business intelligence, 67-82.

[11] Sheakh, Dr.Tariq. (2018). A Study of Inventory Management System Case Study. Journal of Dynamical and Control Systems. 10. 1176-1190.

[12] Plinere, Darya & Borisov, Arkady. (2015). Case Study on Inventory Management Improvement. Information Technology and Management Science. 18. 10.1515/itms-2015-0014.

[13] Anulika, Nwafor & A, Idoko & Agbo, Jonathan & Emeka, Nnaemeka. (2020). Design and Optimization of An Inventory Management System for Central Stores. International Journal of Innovative Research and Development. 05. 119.