IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

OPTIMIZING SUPPLY CHAIN PERFORMANCE WITH SNOWFLAKE DATA WAREHOUSE

[1]Rupesh Neve, [2]Dr. Amol Zade

[1]Research Scholar, [2]Assistant professor

[1]Data Science, G. H. Raisoni University Amravati, India

[2]Data Science, G. H. Raisoni University Amravati, India

Abstract: Supply chain performance plays a critical role in the success of businesses across industries. In recent years, the rise of cloud-based data warehousing platforms has introduced new opportunities for organizations to improve supply chain operations through better data management and analysis. This current dissertation work examines the use of Snowflake Data Warehouse for optimizing supply chain performance. Through a review of literature and analysis of case studies, the paper presents the key features and benefits of Snowflake including cloud technologies, and discusses how it can be integrated with other systems and technologies to support supply chain decision-making and drive performance improvements. The results of the study suggest that Snowflake can be a valuable tool for improving supply chain performance, and offer insights, recommendations and prediction for organizations/business looking to implement and can be visualize in the form of graphs and charts to support quick decision making to the business users.

Keywords- **Machine learning techniques, Cloud storage, Snowflake data warehouse, Visualizations, Data Analysis**.

1. INTRODUCTION

In today's global and highly competitive business environment, supply chain management plays a crucial role in the success of organizations. To maintain a competitive edge and meet customer demands, it is essential to coordinate and optimize the flow of goods and services from raw material suppliers to end customers, which is what effective supply chain management entails. This requires a deep understanding of the various components of the supply chain, such as sourcing, production, logistics, and distribution, as well as the ability to manage and analyze vast amounts of data in order to make informed decisions. The rise of big data and advanced analytics majorly including cloud based data warehouse and cloud based storage services has introduced new opportunities for organizations to improve supply chain performance through better data management and analysis. One platform that has gained significant traction in the supply chain industry is Snowflake Data Warehouse, a cloud based data warehousing solution that offers a range of features and tools for managing and analyzing large volumes of data. In this dissertation work, we will explore the use of Snowflake for optimizing supply chain performance, and present case studies and examples of how the platform has been used in the industry and showcase the various visuals to make strong decisions. We will also discuss the potential benefits and challenges of using Snowflake in the supply chain, and provide insights and recommendations for organizations looking to implement the platform in their operations. The term "cloud" refers to a type of technology that allows users to store and access data and applications over the internet, rather than on their own personal computer or local server. The benefits of using the cloud include increased efficiency, as it allows for shared resources and cost savings, and scalability, which allows users to quickly adjust their usage and storage needs. Additionally, cloud services are often provided on a pay-as-you-go basis, meaning users only pay for the resources they use, rather than having to make a large upfront investment. The traditional data storage service falls short of the demand, because of the massive amount of raw data, the irrelevance between decision-making data from different departments, and the poor query ability of large-scale datasets.

Data warehouses are critical to the world of business intelligence as they gather information from across an entire organization and focus on business processes across the board. In order to support more complex and in-depth analytics, another environment called Online Analytical Processing (OLAP) was created. This system allows users to access the contents of the data warehouse

in a more optimized and structured way, including strategic information to help with decision-making. OLAP supports multidimensional data analysis, which is focused on the relationship between different areas of the business. It maps data elements from the data warehouse to its own data structure to provide users with an advanced yet user-friendly interface.

1.1. DATA WAREHOUSE

Databases accumulate more and more data, leading to an increase in their size. This can create challenges for managing and analyzing the data effectively. To address these challenges, organizations often turn to data warehouses and DataMart, which provide a centralized repository for storing and managing large volumes of data in a way that facilitates analysis and decision-making. A data warehouse is the collection of many databases that are combined for the analysis purpose via quantitatively or qualitatively. DWs like a database that uses the information for reports in a different way to establish new information from previous. DWs is not a full database also can be the specific relevant data that is beneficial for our industry or company. DWs holds the raw data that will not be part of our research.

1.2. Snowflake Data Warehouse

Snowflake Data Warehouse is a cloud-based data warehousing platform designed to allow businesses to store, manage, and analyze their data easily and efficiently. Snowflake was founded in 2012 and is now a publicly traded company. It has quickly become one of the most popular data warehousing platforms on the market due to its unique architecture and advanced features. The Snowflake Data Warehouse is built on top of the cloud, which means that it can handle large amounts of data without requiring on-premises hardware. This allows businesses to scale their data storage and processing capabilities quickly and easily. The Snowflake Data Warehouse also supports a variety of data types and formats, including structured, semi-structured, and unstructured data.One of the key features of the Snowflake Data Warehouse is its unique architecture. Snowflake uses a multi-cluster, shared data architecture that separates compute and storage. This allows businesses to scale their compute resources independently of their storage resources, which is particularly useful for handling fluctuating workloads. The shared data architecture also ensures that data is always consistent across all clusters, regardless of which cluster is being used for processing.Another important feature of the Snowflake Data Warehouse is its support for SQL. Snowflake uses a dialect of SQL that is optimized for cloud-based data warehousing. This allows businesses to leverage their existing SQL skills and tools when working with Snowflake, making it easier to get started with the platform.Snowflake also offers a number of advanced features that make it a powerful data warehousing platform. For example, Snowflake has built-in support for machine learning, which allows businesses to build predictive models and make data-driven decisions. Snowflake also offers support for data sharing, which makes it easy to collaborate with other businesses or departments. Additionally, Snowflake offers a variety of security features, such as data encryption and access controls, to keep data safe and secure
.

1.3. AIM AND OBJECTIVE

The objectives of the topic "A Review on Parental Involvement as Well as its Impact on Academic Achievement" can be converted into paragraph as follows:The aim of this topic is to investigate the role of parental involvement in promoting academic achievement. To achieve this, the first objective is to thoroughly examine the features and capabilities of Snowflake Data Warehouse and its potential benefits for supply chain performance. This involves a detailed analysis of Snowflake's key features such as its ability to handle diverse data types and formats, real-time analytics support, scalability, and flexibility.The second objective is to explore other cloud technologies for storing data using data storage services and suitable storing strategies like Data lake or Lake house. The third objective is to present various case studies and examples of how Snowflake has been used in the supply chain industry. The focus is on analyzing real-world case studies and showcasing the impact of Snowflake on supply chain operations using compelling visuals.The fourth objective is to develop a robust Machine Learning model based on the gathered dataset and train the model with the appropriate dataset to predict supply and demand accurately in the industry. Finally, the fifth objective is to provide valuable insights and recommendations based on the data visuals using visualization tools like Tableau, Power BI or data app websites. This objective focuses on offering insights and predictions to business users looking to make informed decisions for their business or organizations.

2. LITERATURE REVIEW

Benoit Dageville et.al (2016) conducted research on "The Snowflake Elastic Data Warehouse"We live in the golden age of distributed computing. Cloud platforms today provide access to almost limitless computing and storage resources, and the Software-as-a-Service (SaaS) model allows users who previously couldn't afford high-end systems to access enterprise-class technology. However, traditional data warehousing systems are struggling to adapt to this new environment. These systems were designed for fixed resources and are not able to take advantage of the cloud's flexibility. Additionally, their reliance on complex

ETL pipelines and physical tuning conflicts with the cloud's requirement for flexible and constantly updated semi-structured data and workloads. Therefore, we have decided that a complete redesign is necessary. Our mission was to build an enterprise-ready data warehousing solution for the cloud. The result is the Snowflake Elastic Data Warehouse, or "Snowflake" for short. Snowflake is a multi-tenant, transactional, secure, extremely elastic system with built-in extensions for semi-structured and schema-less data. It also fully supports SQL. The system is made available in the Amazon cloud as a pay-as-you-go service. Users can quickly manage and query their data using well-known tools and interfaces after uploading it to the cloud. Installation began in late 2012 and Snowflake has been generally available since June 2015. A rising number of small and large businesses use Snowflake in production today. Across several petabytes of data, the system processes several million requests every day. In this article, we outline Snowflake's innovative multi-cluster, shared-data architecture. The paper emphasises some of the core aspects of Snowflake: extreme flexibility and availability, semi-structured and schema-less data, time travel, and end-to-end security. Lessons learned and an outlook on current work are presented at the end.

Jiangping Wang et.al (2015) conducted research on "Data Warehouse Snowflake Design and Performance Considerations in Business Analytics"Snowflake is a data warehouse schema design where dimension tables are normalized on top of a star schema design. Due to its performance overhead when combining the normalised dimension tables, snowflake schema is generally not advised. . Snows. Snow Snow the Snow Snow Snow improve Snow Snow Snow Snow Snow Snow. Snow Snow the Snow the Snow. Snow Snow Snow the Snow . In business analytics paradigm, two unique environments are complementary and work together to produce successful business analytics. Firstly, the data warehouse environment transforms operational data into information. Second, the analytical environment provides information to users so they may conduct additional data analysis and make decisions. The gap between the two environments is filled by the snowflake schema. Wide dimension structures with several dimension attributes can be mapped to analytical processing hierarchies more easily using the snowflake schema. The snowflake schema makes navigation along hierarchies easier and supports flexible analysis such as drilldown and rollup. This paper examines the two complementary business intelligence environments, roles played by the snowflake design in mapping from data warehouse to analytics, and performance considerations in snowflake design with case studies.

Michael Armbrust et.al (2021) conducted research on "Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics" This paper makes the case that the current data warehouse architecture will eventually become obsolete and be replaced by a new architectural pattern called the Lakehouse. The Lakehouse will I be built on open direct-access data formats like Apache Parquet, (ii) have top-notch support for machine learning and data science, and (iii) provide cutting-edge performance. Data staleness, reliability, total cost of ownership, data lock-in, and limited use-case support are just a few of the significant issues that lake homes may help with**.** We discuss how the industry is already moving toward Lake houses and how this shift may affect work in data management. We also report results from a Lakehouse system using Parquet that is competitive with popular cloud data warehouses on TPC-DS.

YINGYING TAO ET.AL (2003) CONDUCTED RESEARCH ON "OPTIMIZING LARGE STAR-SCHEMA QUERIES WITH SNOWFLAKES VIA HEURISTIC-BASED QUERY REWRITING."

Query optimizers in existing database management systems often suffer from intolerably long optimization time and/or poor optimization results when optimizing large join queries. Before submitting a user-specified complex query to the next query optimisation step, one potential solution to these issues is to rewrite it into a different form that can better leverage the capabilities of the underlying query optimizer, depending on some heuristic principles. We concentrate on researching a specific class of complex query known as a "snow-schema query," which has a star-schema structure and snowflakes. The key idea is to split a given snow-schema query into several levels of small query blocks at the query rewriting stage. The query optimizer then optimizes the query blocks and integrates their results into the final query result. A set of heuristic rules on how to divide the query is introduced. A query rewriting framework adopting these heuristics is presented. Experimental results demonstrate that this heuristic-based query rewriting technique is quite promising in optimizing large snow-schema queries.

Wang Jianmin et.al (2019) conducted research on "An improved join-free snowflake schema for ETL and OLAP of data warehouse"The emergence of big data makes more and more enterprise change data management strategy, from simple data storage to OLAP query analysis; meanwhile, NoSQL-based data warehouse receives more increasing attention than traditional SQL-based database. This paper suggests the uniform distribution code (UDC), model identification code (MIC), standard dimension code (SDC), and attribute dimensional code (ADC), improves the JFSS model for ETL, defines the data storage format, and identifies the extraction, transformation, and loading strategies of data warehouse. Several experiments are carried out to analyze single record and range record queries as typical OLAP based on Hadoop database (HBase). The results show the proposed scheme can provide lower overhead than the traditional SQL-based database while facilitating the scope and flexibility of data warehouse services.Currently, most enterprise management apps only help with data retrieval, collecting, and querying; they fail to understand the value of the enormous volume of historical data and obscure the underlying significance of big data. Enterprise decision-makers are in critical need of the information extrapolated from historical data in the wake of the increasingly fierce market rivalry. In order to meet the demand, practical solutions must be created to extract the necessary data from sizable datasets and determine the relationships between data from multiple departments.

Geetika Saxena et.al (2014) conducted research on "Data Warehouse Designing: Dimensional Modelling and E-R Modelling" The Data Warehouse (DW) is regarded as a collection of historical data that has been consolidated, detailed, and gathered from many sources. Data that will support managerial decision-making is gathered via DW. The conceptual and logical design phases of a data warehouse can be approached in a variety of ways. Dimensional fact models, multidimensional E/R models, starER models, and object-oriented multidimensional models are the conceptual design techniques. Also, there are logical design techniques such as flat schema, star schema, fact constellation schema, galaxy schema, and snowflake schema. We compared both dimensional modelling and E-R modelling in the data warehouse in this essay. The most common data warehousing technique is dimensional modelling (DM). In DM, the decision support query is optimised using a model of tables and relations. Performance of queries in relational databases. Also, traditional E-R models are used to reduce data model redundancy, make it easier to retrieve certain records with specific crucial IDs, and improve the efficiency of online transaction processing (OLTP).

Songting Chen et.al (2010) conducted research on "Cheetah: A High Performance, Custom Data Warehouse on Top of MapReduce" Large-scale data analysis has become increasingly important for many enterprises. Due to its amazing scalability and flexibility to handle both structured and unstructured data, a new distributed computing paradigm known as MapReduce and its open source implementation Hadoop have recently gained widespread use. In this work, we present our MapReduce-based data warehouse system, dubbed Cheetah. Cheetah was created expressly to accommodate different simplifications and personalised optimisations for our online advertising application. First, we take a fresh look at the data warehouse schema design. In particular, we define a virtual view on top of the common star or snowflake data warehouse schema. This virtual view abstraction not only allows us to design a SQL-like but much more succinct query language, but also makes it easier to support many advanced query processing features. We then go over a variety of optimisation methods, including data compression, access methods, multi-query optimisation, and taking advantage of materialised views. In reality, our cluster's commodity hardware enables each node to analyse raw data at a rate of 1 GBytes/s. Then, we demonstrate how to incorporate Cheetah into any ad hoc MapReduce jobs. This allows MapReduce developers to fully leverage the power of both MapReduce and data warehouse technologies.

Zhang Xiaotao et.al (2018) conducted research on "Design of Intelligent Management Decision Support System for Retailing Chains" We analyze data source, data warehouse structure design, OLAP system structure and front-end display tool, and realize system prototype through historical data. Then, data warehouse technology is used to organize and store chain supermarket supply-chain to manage data warehouse and sets up data warehouse technology , OLAP analysis technology-based supply chain performance system and supply-chain management decision support function. The system performs multi-angle analysis in supply-chain, provides management with highly comprehensive information which reflects inventory, retailing and order tendency in order to improve scientific and information degree in decision making.

Rendy Renaldi Berahim et.al (2021) conducted research on "Analysis and Design Optimize Data for the Depok Center Information System's Health Services on the E- Government Data Warehouse Application using an Olap Pivot Table and Star Schema" Interoperability is a hot topic for institutions and researchers worldwide thanks to the advent of new technology, equipment, and communication Suggestions. The amount of data integrated is now increasing year by year, month by day even. So, traditional algorithms and technologies are now inefficient in processing, analyzing, and storing vast amounts of data. Increased use of data integration leads to more efficient technologies urgently needed to divide data and process data between data sources. Interoperation of information system applications into the urgent requirement for e-gov development in Indonesia today. The ultimate goal to achieve through the interoperable solution between the e-gov applications is the creation of an integrated and communicable information system. In the study, data management warehouse could drive the quality of monitoring of leaders faster and more optimally. So it can be used as information and strategic services that make it easier for the public to access information and services provided by the city's Depok government. And for monitoring the activity of each device area can also be monitored by leaders to assist performance assessments of each center as a generated data contributor to and performance so that data accuracy and performance on the application increases.

C.K.M. Lee et.al (2011) conducted research on "Design and development of logistics workflow systems for demand management with RFID" This paper discusses demand and supply chain management and examines how artificial intelligence techniques and RFID technology can enhance the responsiveness of the logistics workflow. This proposed system is expected to have a significant impact on the performance of logistics networks by virtue of its capabilities to adapt unexpected supply and demand changes in the volatile marketplace with the unique feature of responsiveness with the advanced technology, Radio Frequency Identification (RFID). Recent studies have found that RFID and artificial intelligence techniques drive the development of total solution in logistics industry. Apart from tracking the movement of the goods, RFID is able to play an important role to reflect the inventory level of various distribution areas. In today's globalized industrial environment, the physical logistics operations and the associated flow of information are the essential elements for companies to realize an efficient logistics workflow scenario. Basically, a flexible logistics workflow, which is characterized by its fast responsiveness in dealing with customer requirements through the integration of various value chain activities, is fundamental to leverage business performance of enterprises. The significance of this research is the demonstration of the synergy of using a combination of advanced technologies to form an integrated system that helps achieve lean and agile logistics workflow.

Shastri L Nimmagadda et.al (2019) conducted research on "On Modelling Big Data Guided Supply Chains in Knowledge-Base Geographic Information Systems" management systems . Modelling supply chain management systems is held back because of lack of consistent and poorly aligned data with supply chain elements and processes. The issues constraining the decision-making process limit the connectivity between supply chains and geographically controlled database systems. The heterogeneous and unstructured data are added challenges to connectivity and integration processes. The research focus is on analysing the data heterogeneity and multidimensionality relevant to supply chain systems and geographically controlled databases. In pursuance of the challenges, a unified methodological framework is designed with data structuring, data warehousing and mining, visualization and interpretation artefacts to support connectivity and integration process. Multidimensional ontologies, ecosystem conceptualization and Big Data novelty are added motivations, facilitating the relationships between events of supply chain operations. The models construed for optimizing the resources are analysed in terms of effectiveness of the integrated framework articulations in global supply chains that obey laws of geography. The integrated articulations analyzed with laws of geography can affect the operational costs, sure for better with reduced lead times and enhanced stock management.

Daniel L. Moody et.al (2000) conducted research on "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design" This paper describes a method for developing dimensional models from traditional Entity Relationship models. This can be used to design data warehouses and data marts based on enterprise data models. The first step of the method involves classifying entities in the data model into a number of categories. The second step involves identifying hierarchies that exist in the model. The final step involves collapsing these hierarchies and aggregating transaction data to form dimensional models. A number of design alternatives are presented, including a flat schema, a terraced schema, a star schema and a snowflake schema. We also define a new type of schema called a star cluster schema. This is a restricted form of snowflake schema, which minimizes the number of tables while avoiding overlap between different dimensional hierarchies.

Midhul Vuppalapati et.al (2020) conducted research on "Building an Elastic Query Engine on Disaggregated Storage"We present operational experience running Snowflake, a cloud based data warehousing system with SQL support similar to state-of-the-art databases. Snowflake design is motivated by three goals: (1) compute and storage elasticity; (2) support for multi-tenancy; and, (3) high performance. Over the last few years, Snowflake has grown to serve thousands of customers executing millions of queries on petabytes of data every day. This paper presents Snowflake design and implementation, along with a discussion on how recent changes in cloud infrastructure (emerging hardware, fine-grained billing, etc.) have altered the many assumptions that guided the design and optimization of Snowflake system. Using data collected from various components of our system during execution of 70 million queries over a 14-day period, our study both deepens the understanding of existing problems and highlights new research challenges along a multitude of dimensions including design of storage systems and high-performance query execution engines.

MUHAMMAD ZAFAR IQBAL KARMANI ET.AL (2020) CONDUCTED RESEARCH ON "A REVIEW OF STAR SCHEMA AND SNOWFLAKES SCHEMA" IN THE NEW AGE, DIGITAL DATA IS THE MOST IMPORTANT SOURCE OF ACQUIRING KNOWLEDGE. FOR THIS PURPOSE, COLLECT DATA FROM VARIOUS SOURCES LIKE WEBSITES, BLOGS, WEBPAGES, AND MOST IMPORTANT DATABASES. DATABASE AND RELATIONAL DATABASES BOTH PROVIDE HELP TO DECISION MAKING IN THE FUTURE WORK. NOWADAYS THESE APPROACHES BECOME TIME AND RESOURCE CONSUMING THERE FOR NEW CONCEPT USE NAME DATA WAREHOUSE. WHICH CAN ANALYZE MANY DATABASES AT A TIME ON A COMMON PLATE FROM WITH VERY EFFICIENT WAY. IN THIS PAPER, WE WILL DISCUSS THE DATABASE AND MIGRATION FROM THE DATABASE TO THE DATA WAREHOUSE. DATA WAREHOUSE (DW) IS THE SPECIAL TYPE OF A DATABASE THAT STORES A LARGE AMOUNT OF DATA. DW SCHEMAS ORGANIZE DATA IN TWO WAYS IN WHICH STAR SCHEMA AND SNOWFLAKES SCHEMA. FACT AND DIMENSION TABLES ORGANIZE IN THEM. DISTINGUISHED BY NORMALIZATION OF TABLES. NATURE OF DATA LEADS THE DESIGNER TO FOLLOW THE DW SCHEMAS ON THE BASE OF DATA, TIME AND RESOURCES FACTOR. BOTH DESIGN-MODELING TECHNIQUES COMPARE WITH THE EXPERIMENT ON THE SAME DATA AND RESULTS OF APPLYING THE SAME QUERY ON THEM. AFTER THE PERFORMANCE EVALUATION, USING BITMAP INDEXING TO IMPROVE THE SCHEMAS PERFORMANCE. WE ALSO PRESENT THE DESIGN MODELING TECHNIQUES WITH RESPECT TO DATA MINING AND IMPROVE QUERY OPTIMIZATION TECHNIQUE TO SAVE TIME AND RESOURCE IN THE ANALYSIS OF DATA.

Mark Levene et.al (2003) conducted research on "Why is the snowflake schema a good data warehouse design?" Database design for data warehouses is based on the notion of the snowflake schema and its important special case, the star schema. The snowflake schema represents a dimensional model which is composed of a central fact table and a set of constituent dimension tables which can be further broken up into sub dimension tables. We formalize the concept of a snowflake schema in terms of an acyclic database schema whose join tree satisfies certain structural properties. We then define a normal form for snowflake schemas which captures its intuitive meaning with respect to a set of functional and inclusion dependencies. We show that snowflake schemas in this normal form are independent as well as separable when the relation schemas are pairwise incomparable. This implies that relations in the data warehouse can be updated independently of each other as long as referential integrity is maintained. In addition, we show that a data warehouse in snowflake normal form can be queried by joining the

relation over the fact table with the relations over its dimension and sub dimension tables. We also examine an information-theoretic interpretation of the snowflake schema and show that the redundancy of the primary key of the fact table is zero.

Xiaoyan Wan et.al (2011) conducted research on "Application Research of Pharmaceutical Industry Management System Based on ERP and CRM Integration" Design of a pharmaceutical industry management system based on ERP and CRM integration, including design background, function and implement, are presented in this paper. The designed system applies middleware technology in data exchange between ERP and CRM, and all the data is transferred in XML format. The mainly integrated function modules are customer management module, product information module and sales module. Snowflake mode is used in data warehouse to eliminate data redundancy and to save storage space. The system can improve management ability and competition superiority of the pharmaceutical industry through ERP and CRM integration.To obtain survival and development, Enterprise's goal is to maximize the benefits. Efficiency and effectiveness is a key measure of enterprise value, while improving the overall management and management efficiency is the key means to achieve this goal. The integration of the internal resources with ERP as the core and the external resources with CRM as the core provides the tools to achieve the target.

Per-Åke Larson et.al (2011) conducted research on "SQL Server Column Store Indexes" The SQL Server 11 release (code named "Denali") introduces a new data warehouse query acceleration feature based on a new index type called a column store index. The new index type combined with new query operators processing batches of rows greatly improves data warehouse query performance: in some cases, by hundreds of times and routinely a tenfold speedup for a broad range of decision support queries. Column store indexes are fully integrated with the rest of the system, including query processing and optimization. This paper gives an overview of the design and implementation of column store indexes including enhancements to query processing and query optimization to take full advantage of the new indexes. The resulting performance improvements are illustrated by a number of example queries.In this case, a column-wise organization where values from the same column in different records are stored together performs much better. It reduces the data processed by a query because the query reads only the columns that it needs and, furthermore, column-wise data can be compressed efficiently. Systems using column-wise storage are usually referred to as column stores.

3. PROPOSED SYSTEM

In order to provide a comprehensive solution for supply and demand analysis, we have developed a two-pronged approach. The first approach involves providing end-to-end analytics, which involves analyzing the entire supply chain from raw materials to finished products and identifying bottlenecks and areas for improvement. The second approach involves using predictive modeling techniques to forecast future demand and supply, allowing businesses to proactively plan and adjust their operations to meet changing market conditions. By combining these two approaches, we aim to provide a comprehensive and robust solution for supply and demand analysis.

Analytics: In order to optimize transportation, our analytics approach involves analyzing the entire supply chain from raw materials to finished products and identifying inefficiencies and bottlenecks. By doing so, we can determine the best route for transporting raw materials from refineries to warehouses and from warehouses to retail outlets, minimizing transportation costs and ensuring timely delivery of goods. Additionally, our analytics can help identify opportunities for reducing waste by finding ways to repurpose or recycle raw materials. Finally, by analyzing demand patterns, we can optimize the use of warehouses to ensure that they are not overstocked or under stocked, reducing the risk of waste and improving inventory management. Overall, our analytics approach aims to provide a comprehensive and cost-effective solution for optimizing transportation and supply chain management.

Prediction: In order to predict demand based on supply, we can use machine learning algorithms to forecast future demand and identify trends and patterns. By training the algorithm on historical data, it can learn to make accurate predictions about future demand based on a variety of factors, including supply levels, market conditions, and consumer behavior. Additionally, we can use feedback and sentiment analysis to gather insights from customers and improve the accuracy of our demand forecasts. By combining machine learning with feedback and sentiment analysis, we can provide a 13 more comprehensive and accurate prediction of future demand based on current supply levels. This approach can help businesses proactively plan and adjust their operations to meet changing market conditions and avoid shortages or excess inventory.

3.1. Business Flow

Ideal business flow for our research topic would be similar to this. As we have proposed the approaches in the solution (above point) we analyze the transportation, warehouse and wastage to optimize supply chain.

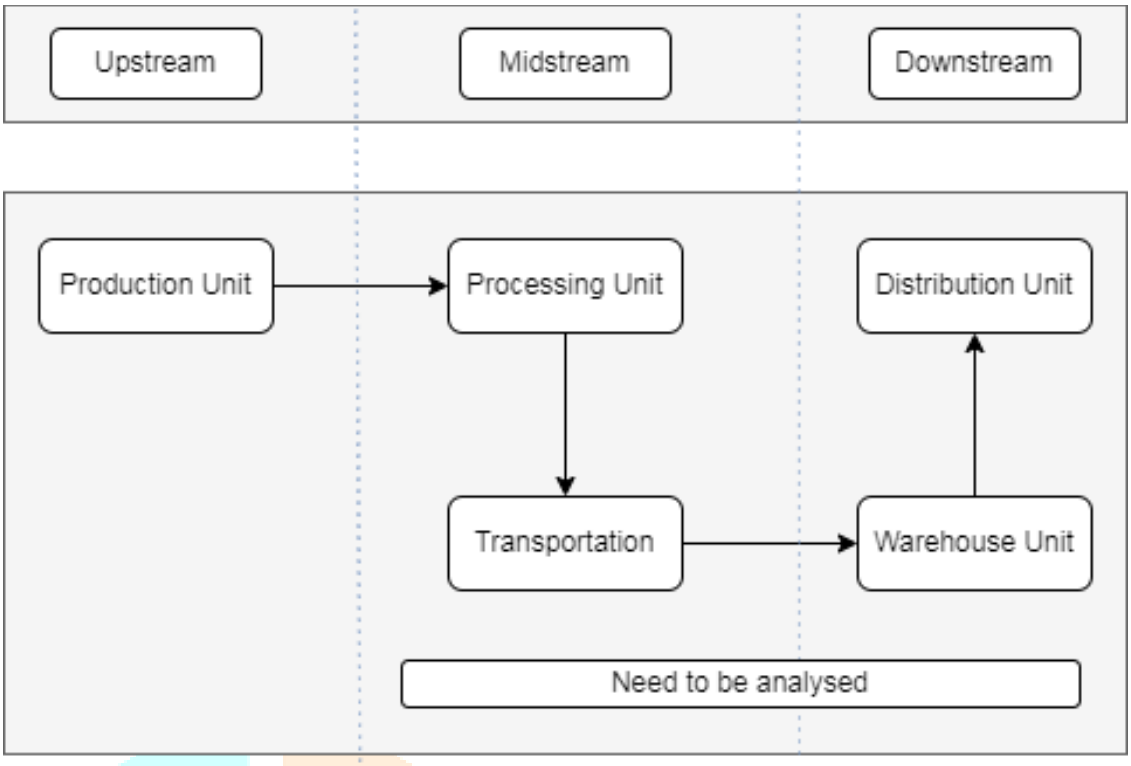

Fig.1. Business Flow

3.2. Architecture

Snowflake is a data warehousing tool that allows businesses to store and manage large amounts of structured and unstructured data. The architecture of Snowflake typically involves the use of various data sources, which may include transactional systems, operational databases, and external data sources such as social media and web APIs. These data sources feed into the Snowflake data warehouse, where the data is cleansed, transformed, and organized for efficient storage and analysis. 14 Architecture is designed to provide businesses with a powerful, scalable, and secure platform for managing and analyzing data, enabling them to make better decisions and drive business growth.

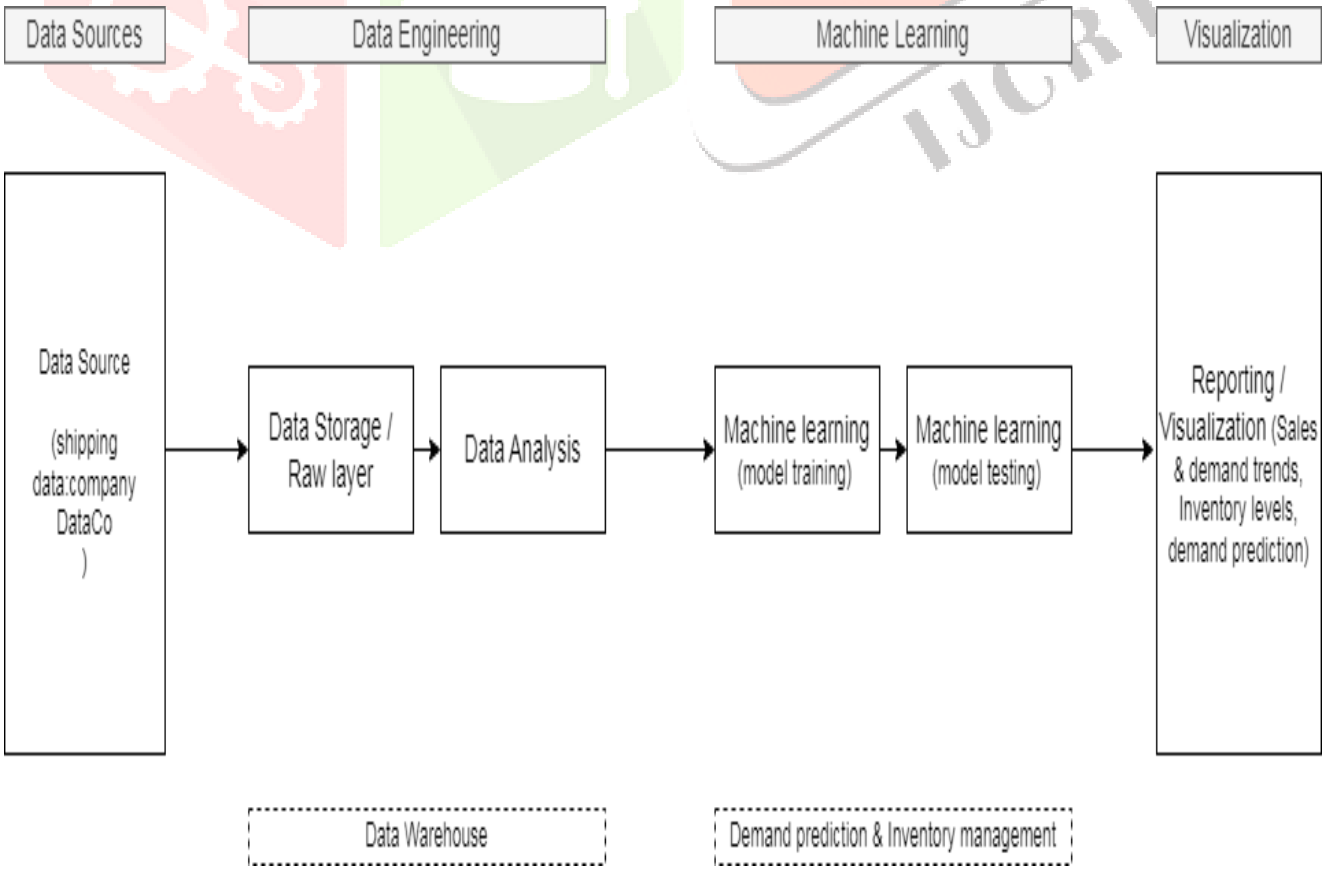

Fig. 2. Architecture

4. RESULTS AND OUTCOMES

The potential results and outcomes of the dissertation for supply chain optimization are significant and can have far-reaching impacts on a business's operations, customer satisfaction, and bottom line.

- Improved data accessibility and visibility: Snowflake Data Warehouse provides a centralized platform that allows supply chain stakeholders to access and analyze data in real-time, which can lead to faster and more informed decision-making.
- Enhanced operational efficiency: By providing real-time data analytics and predictive modeling capabilities, Snowflake Data Warehouse can help businesses identify inefficiencies and opportunities for improvement in their supply chain operations, leading to increased operational efficiency and cost savings.
- Better demand planning: With the ability to handle diverse data types and support real-time analytics, Snowflake Data Warehouse can help businesses gain insights into customer demand and plan inventory levels more accurately, resulting in reduced stockouts and overstocking.
- Improved inventory management: By building a strong Machine Learning model for predicting supply and demand, businesses can optimize inventory levels, reduce waste, and improve order fulfillment rates.

Suggestions for businesses contemplating using Snowflake data warehouse for supply chain optimization:

We propose that firms adopting Snowflake data warehouse for supply chain optimization thoroughly evaluate the platform's capabilities and possible effect on operations. Data integration, security, scalability, and cost-effectiveness should all be factors to consider. Businesses should also examine the advantages and disadvantages of utilising sophisticated analytics and machine learning in supply chain management. Overall, Snowflake data warehouse is a tempting alternative for businesses looking to improve supply chain performance and acquire a competitive advantage in their market.

The application of ML algorithms along with snowflake datawarehouse in various aspects of supply chain management has revolutionized the way businesses operate by enabling them to optimize their operations, reduce costs, and improve customer satisfaction.

Demand forecasting plays a pivotal role in supply chain planning as it enables businesses to anticipate future demand for their products or services. This is achieved through the use of various machine learning (ML) algorithms such as neural networks that analyze historical sales data and other relevant factors to predict future demand accurately. Among these, Artificial Neural Networks (ANNs) with relative error of less than 2% have proven to be more accurate than linear regression models (relative error of 16-20%) in demand forecasting.

Inventory optimization is another area where ML algorithms, specifically ANNs, can be used to improve supply chain efficiency. By identifying demand patterns and predicting future inventory requirements, businesses can optimize their inventory levels to avoid overstocking or stockouts, ultimately leading to a reduction in operational costs.

Transportation optimization is yet another area where ML algorithms such as linear regression and genetic algorithms can be used to analyze historical data and identify optimal routes for transportation, resulting in reduced transportation costs and more efficient logistics. ML algorithms such as reinforcement learning and deep learning can be applied to optimize warehouse management by improving inventory management, reducing fulfillment times, and optimizing warehouse layouts. This is achieved by analyzing various factors such as inventory turnover rate, order volume, and storage capacity to develop efficient and effective warehouse management strategies.

5. CONCLUSION

Snowflake Data Warehouse is a powerful tool for managing and analyzing complex supply chain data. It provides a centralized platform that allows supply chain stakeholders to access and analyze data in real-time, enabling better decision-making and improved operational efficiency. Cloud-based data storage services such as Data Lake or Lake House are popular options due to their ability to store large volumes of data in a scalable and cost-effective manner. To maximize the benefits of these data storage services, appropriate storing strategies need to be implemented, such as data lifecycle management, data encryption, and access controls. Snowflake has proven to be a powerful tool in the supply chain industry, with numerous case studies demonstrating its effectiveness. It can provide supply chain stakeholders with the necessary insights to make informed decisions, improve operational efficiency, and ultimately drive business success.

The impact of Snowflake on the supply chain industry can be seen through its ability to handle diverse data types and formats, support real-time analytics, and provide scalability and flexibility. Additionally, the use of impactful visualizations allows businesses to quickly identify patterns, trends, and anomalies, leading to faster decision-making and improved operational efficiency. Building a strong prediction Machine Learning model for predicting supply and demand in the industry is a crucial step towards achieving operational efficiency and business success. The key to building a strong model for supply and demand prediction lies in the quality of the training dataset, which must be diverse, representative, and properly pre-processed. Additionally, businesses need to use appropriate Machine Learning algorithms and techniques, such as linear regression, decision trees, or neural networks, to build a robust and accurate model. The impact of this model on supply and demand can be far-reaching, including improved inventory management, better demand planning, and enhanced operational efficiency.

Building a strong prediction Machine Learning model for supply and demand is a critical step towards achieving operational efficiency and business success. Visualization tools such as Tableau, Power BI, or data app websites have become increasingly important in providing insights and recommendations to businesses looking to make informed decisions. These tools can help businesses gain a better understanding of their data and identify key patterns and trends, leading to more accurate predictions and informed decision-making. Additionally, they can help businesses identify areas for improvement and develop strategies for addressing them. The use of visualization tools is not limited to large businesses alone, but small and medium-sized businesses can also benefit from the insights and recommendations offered by these tools. With the increasing availability of cloud-based visualization tools, businesses of all sizes can leverage these technologies to gain a competitive edge in their respective industries.
The potential of Snowflake data warehouse in enhancing supply chain performance was investigated in this seminar report. We've spoken about the benefits of Snowflake data warehouses, key performance indicators for supply chain performance, and the obstacles of improving supply chain performance. We also looked at how Snowflake data warehouse can be utilised for real-time data integration, data analytics and visualisation, as well as machine learning and artificial intelligence. The case studies given show the potential advantages of adopting Snowflake data warehouse for supply chain optimization.
The ramifications of adopting Snowflake data warehouse to optimise supply chains are enormous. Companies may achieve a competitive advantage in the fast-paced and complicated world of supply chain management by employing real-time data integration, sophisticated analytics and visualisation, machine learning and artificial intelligence. Snowflake data warehouse is a powerful tool for businesses looking to enhance supply chain efficiency, save costs, and boost customer happiness.

REFERENCES

[1] Dageville, Benoit, et al. “The Snowflake Elastic Data Warehouse.” Proceedings of the ACM SIGMOD International Conference on Management of Data, vol. 26-June-2016, 2016, pp. 215–26, https://doi.org/10.1145/2882903. 2903741.

[2] Wang, Jiangping, and Janet L. Kourik. “Data Warehouse Snowflake Design and Performance Considerations in Business Analytics.” Journal of Advances in Information Technology, vol. 6, no. 4, 2015, pp. 212–16, https://doi.org/10.12720/jait.6.4.212-216.

[3] Armbrust, Michael, et al. “Lakehouse: A New Generation of Open Platforms That Unify Data Warehousing and Advanced Analytics.” Cidr, 2021.

[4] Tao, Yingying, et al. “Optimizing Large Star-Schema Queries with Snowflakes via Heuristic-Based Query Rewriting.” IBM Centre for Advanced Studies Conference, no. June 2014, 2003, https://doi.org/10.1145/961322.961365.

[5] Jianmin, Wang, et al. “An Improved Join-Free Snowflake Schema for ETL and OLAP of Data Warehouse.” Concurrency and Computation: Practice and Experience, vol. 32, no. 23, 2020, pp. 1–11, https://doi.org/10.1002/cpe.5519.

[6] Saxena, Geetika, and Bharat Bhushan Agarwal. “Data Warehouse Designing: Dimensional Modelling and E-R \nModelling.” International Journal of Engineering Inventions, vol. 3, no. 9, 2014, pp. 28–34, http://www.ijeijournal.com/papers/Vol.3-Iss.9/D03092834.pdf.

[7] Chen,Songting.“Http://Www.Vldb2010.Org/Proceedings/Files/Papers/I08.Pdf.”Vldb2010.Org, pp. 1459–68, http://www.vldb2010.org/proceedings/files/papers/I08.pdf%5Cnpapers2://publication/uuid/B303A92E-D708-4CFE-A726-C8E9D06CBCB5.

[8] Zhang, Xiaotao. “Design of Intelligent Management Decision Support System for Retailing Chains.” Proceedings - 2018 International Conference on Virtual Reality and Intelligent Systems, ICVRIS 2018, 2018, pp. 485–89, https://doi.org/10.1109/ICVRIS.2018.00125.

[9] Berahim, Rendy Renaldi, and Mohammad Iqbal. “"Analysis and Design Optimize Data for the Depok Center Information System’s Health Services on the E-Government Data Warehouse Application Using an Olap Pivot Table Analysis and Design Optimize Data for the Depok Center Information System’s Health Services on the E-Government Data Warehouse Application Using an Olap Pivot Table and Star Schema.” International Research Journal of Advanced Engineering and Science, vol. 6, no. 3, 2021, pp. 218–26.

[10] Lee, C. K. M., et al. "Design and Development of Logistics Workflow Systems for Demand Management with RFID." Expert Systems with Applications, vol. 38, no. 5, 2011, pp. 5428–37, https://doi.org/10.1016/j.eswa.2010.10.012.

[11] Nimmagadda, Shastri L., et al. "On Modelling Big Data Guided Supply Chains in Knowledge-Base Geographic Information Systems." Procedia Computer Science, vol. 159, 2019, pp. 1155–64, https://doi.org/10.1016/j.procs.2019.09.284.

[12] Moody, Daniel, and Mark A. ..Kortink. "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design." Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000), vol. 2000, 2000, pp. 1–12.

[13] Draper, Pete. "Snowflakes." Deconstructing the Elements with 3ds Max 6, 2020, pp. 193–202, https://doi.org/10.4324/9780080503837-25.

[14] Iqbal, M. Zafar, et al. "A Review of Star Schema and Snowflakes Schema." Communications in Computer and Information Science, vol. 1198, no. May, 2020, pp. 129–40, https://doi.org/10.1007/978-981-15-5232-8_12.

[15] Levene, Mark, and George Loizou. "Why Is the Snowflake Schema a Good Data Warehouse Design?" Information Systems, vol. 28, no. 3, 2003, pp. 225–40, https://doi.org/10.1016/S0306-4379(02)00021-2.

[16] Wan, Xiaoyan, and Guoqing Xu. "Application Research of Pharmaceutical Industry Management System Based on ERP and CRM Integration." Applied Mechanics and Materials, vol. 101–102, 2012, pp. 55–59, https://doi.org/10.4028/www.scientific.net/AMM.101-102.55.

[17] Larson, Per Åke, et al. "SQL Server Column Store Indexes." Proceedings of the ACM SIGMOD International Conference on Management of Data, 2011, pp. 1177–84, https://doi.org/10.1145/1989323.1989448.

[18] Ildikó Gyurkóné Balázs and László Duma PhD, "The Data in Supply Chain Management", Budapest Business School, College of Finance and Accountancy, Information Technology Department, Budapest, Hungary Associate professor, Budapest Corvinus University, Info communication Department, Budapest, Hungary.

[19] Ivan Varela Rozados and Benny Tjahjono, "big data analytics in supply chain management: trends and related research", 6th International Conference on Operations and Supply Chain Management, Bali, 2014.

[20] Hui Hu, Guangyu Zhu, Yanjun Bo, Jinsheng Shen, "Study on a Combined Demand Forecasting Model of the Supply Chain", 'Traffic and Transportation School, Beijing Jiaotong University, Beijing, P.R.China.

[21] HongJing Liu, "Forecasting Model of Supply Chain Management Based on Neural Network", Department of Economics,Qinhuangdao Institute of Technology,Qinhuangdao, 066000, China

[22] Praveen K B, Prateek J, Pradyumna Kumar, Pragathi G, Prof. Madhuri J, "Inventory Management using Machine Learning" , International Journal of Engineering Research & Technology (IJERT)