ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 8 Issue 12 December 2019, Page No. 24918-24938
ISSN: 2319-7242 DOI: 10.18535/ijecs/v8i12.4422

Big Data Engineering for Real-Time Inventory Optimization in Wholesale Distribution Networks

Avinash Pamisetty

Mulesoft Developer, ORCID ID: 0009-0002-0253-4623

Abstract

Supply chains represent the next frontier in Big Data. In an increasingly competitive environment, the need for supply chain optimization and real-time supply chain management becomes evident. For proper design and evaluation of forecasts, orders, and other supply chain parameters, comprehensive and timely information is essential. Significant operational improvements can be made by ensuring the proper functioning of inventory flows across the supply chain. Our overarching goal is to address challenges in real-time inventory management in Supply Chain Networks using Data Engineering. A wholesale distribution network with a warehouse and retailers is considered as a case study. We work with data-rich situations, where there is a wealth of information available in the form of incoming orders at the retailers.

In summary, the goal is to estimate demand parameters (lead-time, post-lead time distribution), optimize order schedules (policies) such that order quantities are minimized holding g-costs, and develop infrastructure and deployment solutions for proper implementation of the Case Study. The emphasis lies on parameter estimates. Each retailer places an order according to a policy which is the main focus of this work. Extensive focus is put on the database aspect, which estimates nonlinear delay and time-lag of the up-to-order retailer. A scripting language processes up-to-order information from any retailer and ensures that the coded and prefiltered order data can be easily integrated to different Enterprise Resource Planning Systems or utilized in any commercial MRP-Inventory Management Framework.

Overall, understanding and manipulating the order data can readily lead to significant reductions in behavioral costs associated with orders that are not properly scheduled. Moreover, improved estimates can serve as input parameters in other estimators or optimizers that involve larger groups of retailers and their orders.

Keywords: Real-Time Data Processing,Inventory Optimization,Big Data Analytics,Wholesale Distribution,Supply Chain Visibility,Stream Processing,Demand Forecasting,Data Lake Architecture,Apache Kafka,Predictive Analytics,IoT Integration,Distributed Systems,Data Pipeline Automation,Machine Learning Models,Operational Efficiency.

1. Introduction

Shakeout and Industry Consolidation. The advent of big data and the associated new IoT technologies, along with improvements in sensor technologies, and advances in communication technologies, have unveiled new opportunities in various industries. For instance, RFID, GPS, and other sensors provide detailed inventory data that were previously unavailable. Today's inventory data can take the shape of a time series consisting of a sequence of various time-stamped states. With the innovation in cloud storage, the cost for storing this data has significantly declined, leading to the growth of large-scale data storage. Although the management and control of inventory have been improving industrially, less research work has utilized this raw data on-line and in real-time to improve the efficiency of this operation.

Big Data Engineering for Supply Chain Applications. This endeavor aims to develop real-time data-driven strategies for the inventory operations of a very large-size distribution network comprising a large number of inventory locations or warehouses. It focuses on two key functionalities, data-driven forecasting of on-hand inventories and demand, and data-driven optimization, where the goals are: to propose data-driven state-of-the-art methodologies to market practitioners; to provide detailed explanations of the theoretical materials, numerical methodologies, and computational designs behind these methodologies to both academic and industrial communities; and to implement end-to-end software solutions for research or instructional purposes.

Technical Challenges. The first mission is on data-driven stochastic optimization via the integrative use of big data and non-parametric machine learning for products with time-sensitive demand. More specifically, upon facing with hyperparameters for which parameter richness is often mis-specified, a non-parametric compositional neural network approach is developed to achieve a data-driven near-optimal Holt-Winters' seasonal exponential smoothing estimator for the sum-of-squared-error loss function. To mitigate the legacy inter-dependency challenge of team-based decision-making, a framework is developed for distributed on-screen decision-forecasting-making of large-scale warehouse networks that combines static heuristic modeling with real-time high dimensional hybrid forecasting.

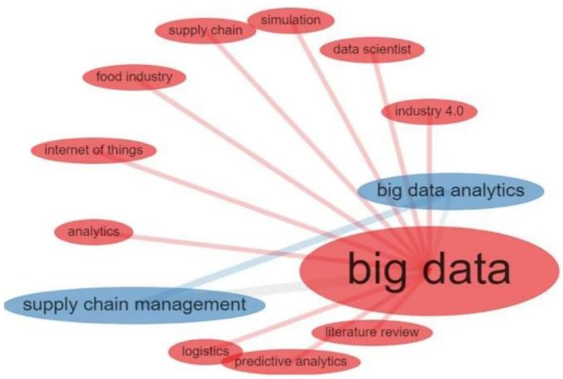


Fig 1: Big data optimisation and management in supply chain management

1.1. Background And Significance

This dissertation focuses on data-driven inventory optimization in wholesale distribution networks. The recent advances in internet technologies, computing devices, and transportation systems have generated massive streams of data, known as Big Data. Useful information can be extracted from these data streams through advanced Big Data technologies . Many issues associated with Big Data have been studied, while discussion of the impacts of these advances on traditional inventory systems is relatively limited. This dissertation fills these gaps and investigates the design of a novel Big Data system for real-time inventory optimization in wholesale distribution networks. Specifically, the research contributions are in the following three aspects:

First, a new massively parallel Big Data system architecture is designed with attention to economic sustainability for real-time forecasting and decision-making in an inventory management environment (I-MT). The design involves equal scaling of application statistical analyses and decision-making software to optimize the total I-MT throughput within a Big Data cloud environment. This development is expected to facilitate the widespread adoption of Big Data I-MTs in organizations of all sizes and industries.

Second, a new I-MT model is designed using stream input generalized seasonal counting (GSC) nets. This advance is expected to bring significant performance improvements for organizations adopting I-MTs, especially those with relatively low IT. The GSC nets allow users to easily manage information overload scenarios that are otherwise difficult to handle under conventional forecasting systems. In cases where over-stored GSC nets become computationally prohibitive, this dissertation designs a new incremental stem GSC (ISGSC) Campbell's theorem to transform the nets into equivalent IS GSC nets.

Last, a new model for real-time inventory allocation and ordering decision-making in wholesale distribution environments is designed. This approach adopts dynamically updated recursive group simulation Schmidt nets, which allow for the tracking of changes in customer demand patterns with minimal performance losses. This improvement is expected to bring significant supply chain performance improvements for organizations adopting Big Data technologies. Formulations and proofs for invariance under partitioned updating are also provided, along with an illustrative example.

Equ : 1 Demand Forecasting using Exponential Smoothing

$$\hat{D}_t = \alpha D_{t-1} + (1-\alpha)\hat{D}_{t-1}$$

- $\hat{D}_t$: forecasted demand at time t
- D_{t-1}: actual demand at time $t-1$
- α: smoothing factor ($0 < \alpha < 1$)

2. Understanding Big Data

The definition of "big" has been changing from megabytes in the 1970s to the petabyte range today. If big data were defined with respect to the amount of data that an organization has to store, process and analyze, every day more organizations would qualify for having big data. Like many other terms, it is defined according to one's perspective and many organizations would be keen to clarify their BDs. BD can be seen as two different issues: on the one hand, big throughput and, on the other, big analytics. The former includes the problems associated with storing and manipulating large amounts of data. The latter are concerned with transforming this data into knowledge.

Focusing on the analytics, BD analytics can be seen as a workflow that distills Terabytes of low-value data down to a single bit of high-value data with the goal to see the big picture. This new discipline needs new approaches to obtain insights from highly detailed, contextualized and rich contents that may require complex math operations, such as machine learning or clustering. Several AI technologies play a crucial role in the analytics of these systems. For instance, top-performing organizations make decisions based on rigorous analysis at more than double the rate of lower performing organizations. Analytic insight is being used to guide both future strategies and day-to-day operations. Literature reports significant interest in the potential of big data and analytics to transform the competitive landscape and improve organizational performance. For instance, examples of the use of big data can be found in government, academia, medicine, climate science and agriculture.

One of the main tools employed in organizations are Decision Support Systems (DSS), which can support complex decision making and problem solving. In the last three decades, there has been considerable technological evolution in both data storage and processing, and in mathematical modeling and analysis. DSS have evolved from sophisticated, stand-alone systems to enterprise wide business intelligence systems that exploit massive amounts of transactional data more than ever before. On the other hand, although a substantial amount of insights from the analysis of business data can be beneficial to the organization, their usage is not apparent, especially at the operational level. Real-time, low latency monitoring and analysis of business events for decision making is key. Such tasks involve problems ranging from the extraction and integration of relevant data from multiple event sources, through the exploitation of embedded knowledge, to the dissemination of monitoring and analytics results to the intended audience. The difficulties are intensified by processes and supply chains that entail dealing with the integration of enterprise execution data across organizational boundaries. The heterogeneity of these supporting systems makes the collection, integration and analysis of high volume business event data, extremely difficult. The new possibilities of storing and analyzing big data are changing the DSS landscape, including decision support social networks.

2.1. Definition of Big Data

It is fair to say that the IT world has been facing challenges for over four decades, but what was “big” in the 1970s has a different meaning today than it does in the petabyte range. It can be seen as two equal but different issues: big throughput and big analytics; the former includes the problems associated with storing and manipulating large amounts of data and the latter those concerned with transforming this data into knowledge. Focused on analytics, big data analytics is defined as a workflow that distills large amounts of low-value data down to a single bit of high-value data with the goal to see the big picture. What separates big data analytics from traditional analytics is the scale; this new discipline requires new approaches in order to obtain insight from highly detailed, contextualized, and rich contents that may require complex math operations, such as machine learning or clustering. This diversity of tools and techniques for big data-driven analytics systems makes the process nontrivial. Complex processes demand complex systems and thus, several artificial intelligence technologies play quite a crucial role.

As mentioned before, very high throughput and very high analytics are needed in order to cope with high-dimensional and high-rate data streams. There has been significant interest in the potential of big data and analytics to transform the competitive landscape and improve organizational performance since the emergence of the big data phenomenon in the early 2000s. Growth in the data itself, advancements in the storage technology, and fast spreading of data servers, sensors, mobile devices along with the well-known growth of social networking sites have contributed significantly to the growth of big data. Examples of the use of big data can be found in several sectors too, including government, academia, medicine, climate science and agriculture. On the other hand, the initial surge of excitement has been replaced by caution and some scepticism. In particular, academics and industry practitioners have voiced concerns regarding privacy issues; while others have pointed out that empirical evidence of the actual value of big data through improved performance remains sparse and questions have been raised on the effectiveness of big data technology, as hindering approaches such as investing in IT or extra data cleansing tend to be more beneficial, especially in highly information dependent industries.

2.2. Characteristics of Big Data

The term ‘big data’ (BD) has a relative meaning. What is big today was not big yesterday, and vice versa. In the 1970s, the megabyte was considered to be a large volume of data. In 2012, the petabyte was viewed as an extraordinarily large volume of data. Clearly, the definition of the ‘big’ is changing over time. However, BD has been abruptly inserted and accepted as a commonplace in government, academia, medicine, climate science and agriculture. There are several things to say about the BD hype. BD can be seen as two different issues: big throughput and big analytics. The management of big throughput includes storing and manipulating a large amount of data, while big analytics transforms this data into knowledge and foresight. BD analytics distills Terabytes of low-value data into a single bit of high-value data to be able to see the big picture. It gives the right insights to the right audience at the right time, so they could decide the right actions to the right targets. This discipline requires new approaches to obtain the insights out of highly detailed, contextualized, and rich contents that may require the most complex math operations, such as pre-cognition, machine learning or clustering. Artificial intelligence technologies play a crucial role in the delivery of these analytics systems, real-time and predictive. Such system-wide surveillance has been increasingly employed for event monitoring and risk management in a variety of application domains. Top-performing organizations make decisions based on rigorous analysis at more than double the rate of lower performing organizations. They are more likely to use analytic insight to guide both future strategies and day-to-day operations. All-winner firms show significant levels of improvement as a direct result of using analytics, with these companies nine times more likely than their competitors to cite analytical insights as an important contributor to their top business impact.

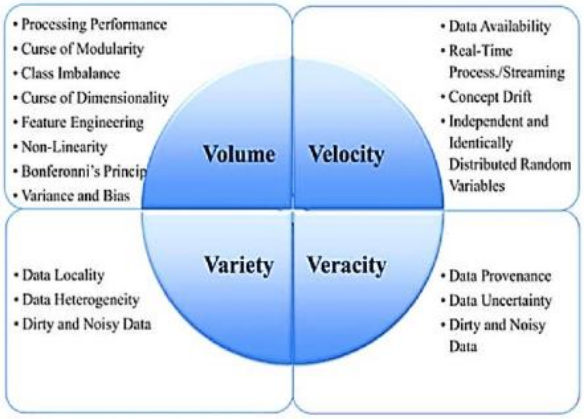


Fig 2: Big data characteristics

2.3. Sources of Big Data in Wholesale Distribution

Based on wide industry experiences, this section presents interviews conducted with five industry practitioners and categorizes the responses of interviewees into several factors, sources, and issues of big data in wholesale distribution or major topics for designing a big data application. The most frequently mentioned big data sources are built-in ERP data of major capabilities, major events from major and 3PL providers, vendors, IoT, and measures for inefficient inventory turnover. The most commonly mentioned applications as big data trend analysis include 1-1.5 year data EDA for distribution center, warehouse analysis for product-line restructuring, and analysis on major stores for promotion effectiveness. The most frequently mentioned data or issue sources for understanding necessary improvements or impacts are customer group & location analysis, model-based advanced forecasting, analysis on ABC group & excess stock, demand & supply research on newly enhanced sales channels, daily demand registration merchandise sample for alarms and understanding data reliability, automatically matched unsold item group based history sales/price analysis, and integrated POS/online sales data. The most frequently mentioned issues are order fulfillment bias between systems, stock-keeping unit data management issues, big data platform development need, and issues with analysis environment, data transformation, and usability. The most frequently mentioned operators to gather necessary terms to help find data sources or design architecture are top-down operators to assist research & analysis or data-based management for strategic thinking. The most frequently mentioned challenges are massively excessive development period, costs of scalable architectures for business impacts, investment priority difference across divisions, intermittent data redundancy across applications, supply chain behavior model calibration needs for architecture development or diagnosis, data person supply or managing issue with know-how loss risk, and delay in detailed behavioral finding analysis for unused knowledge. The most frequently mentioned consultants or supports for modeling and data access are answerable questions and alternative consequences of behavior changes.

3. Importance of Inventory Optimization

In industries dealing with physical materials, management of inventory is very crucial to the success of a company. In wholesale distribution of commodity products, fulfilling demand while managing inventory effectively across warehouses, distribution centers and retailers can lead to significant improvements in operational efficiency, which translates directly into enhanced profit margins. The question at hand is how an organization effectively plans for which products need to be supplied to which retailers, and in what quantities and at what time. This optimization must be done subject to constraints on warehouse capacity, product shelf life, distributor capacity and available truck-fleet. In designing a detailed model on inventory optimization in wholesale drug distribution networks, the supply, storage, shelf-life and demand-side characteristics of the drug industry have been incorporated as an integral part of the model. Development of mathematical formulation and preliminary experiments on simulation and optimization of this model on a small-sized network provide encouraging results.

A better management of inventory can improve a firm's bottom line profit significantly. In the title industry, the distributor companies usually handle large quantities of costly products affecting the profit of the company. These companies have multiple facilities that belong to different sizes. In such facilities, products with different shelf lives are handled. Since most of these products are costly, a better management of a company's inventory can significantly improve its profit. Most of the inventory management systems in the industries are manual. Based on inventory analysis, index number analysis and statistical tools, the decision to replenish inventory and the quantity to reorder are taken. Based on this demand, distributors manually replenish inventory from their suppliers. There are a few industry or research initiatives that focus on the specific inventory decision. Handling large inventory in a facility is also a problem. Large inventory in a warehouse implies high storage and holding costs. Analyzing product profile and arranging a FIFO mechanism will result in the smaller inventory and greater warehouse turnover. A product in a warehouse is replenished when the available quantity goes below the prespecified minimum level. This minimum level depends on the prices of that product in the suppliers, warehouse and retailers.

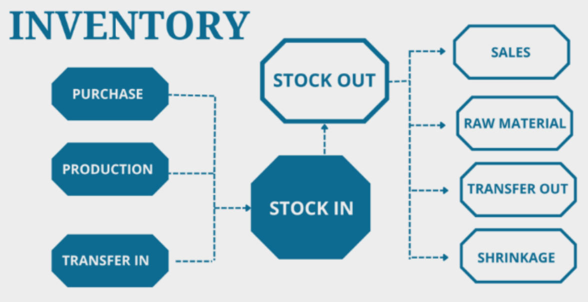


Fig 3: Inventory Optimization

3.1. Challenges in Inventory Management

One of the well-known challenges in inventory management is to characterize the uncertainty of demand. Given its important impact on operation practice, significant advances have been achieved in both parametric and non-parametric models to address the richer variant of the problem. In practice, firms often confront tougher variants of the problem, in which direct implementation of either parametric or non-parametric models is hard. Three common challenges are addressed. First, firms have limited historical data as they are relatively young in the industry. Second, the demand is very low meaning that either it is enough to fill a single truck load or a default positive demand is a highly unlikely event. Finally, as current inventory levels approximate zero, it is hoped to optimally determine inventory targets accounting for a very high service requirement, meaning low stock-out risk.

Setting inventory targets under limited data is a practical challenge for firms, since one of the central tasks in a non-parametric model is to estimate a quantile of underlying distribution with observed data. High service requirements are usually problematic for non-parametric models, since these models need to wait for a long enough time for a small-chance event to show up in observations to account for them. In a hybrid model, to incorporate the availability of demand-simulation policy, it is a non-traditional inverse-reinforcement-learning setting, which models a trader executing a learning-reacting strategy to a market impact at a fixed price, and provides with remarkable improvement for both simulated and real data applications. Quantifying how relief is obtained by continuously reproducing halted or killed transaction trajectory is also of interest.

3.2. Benefits of Real-Time Inventory Optimization

With the emerging data technologies, supply chain networks can be actively optimized similar to stock exchanges, so that delivery routes and the timing and quantity of delivery can be adjusted freely, making these supply chains more sustainable. It is essential to optimize product lifecycle management in supply chain networks since efficiency in parts ordering and delivery considerably affects overall performance. As the depth and scale of outstanding orders have rapidly expanded, balance of delivery schedule and optimization of delivery route become crucial for the performance of supply chain networks. Delivery requests vary due to seasonal product purchasing and production uncertainties of spare parts. It is highly challenging to accurately optimize delivery schedules and simultaneously balance it with outstanding orders in near real-time owing to nonconvex optimization problems, unpredictable demand changes, and enormous solution spaces. Wholesale distribution networks currently manage inventory synchronization by looking back at supply side fuel consumption. However, there are some theoretical risks, including inventory-oblivious external parameters, covertly negotiated fuel cost, delivery grouping among multiple inventory levels, and unprecedented environmental factors. Consequently, there arises a simultaneity challenge on buyer side refill decisions that have to occur when the price is dropping and the seller side demands base stock adjustments that are required before the current state falls into the risk zone. Data engineering technologies allow for the monitoring of the market indicators. In this way, distribution network-wide inventory synchronization can be modeled as a novel class of middle-term, distributed, soft-constraint, real-time extensions of the parametrized timed automata, which require less than 100 parameters to be tuned through the entire centralized scheduling history. A system architecture consisting unattended weather stations to observe supply and demand conditions, data pipeline to filter and carry guarded rules by a core equilibrium-calibrated Q-function approximator, distributed reinforcement learning agents on the unsupervised section of the graph, centralized batch learning agent on the supervised section of the graph, and event-driven task schedules allows for adaptation to severe environmental uncertainty. Rigorous mathematical formulations and effective heuristics of average temperature matrix factorization provide approximated but real-time solutions. Response surface methods capture the invariant of nonlinear kernelized relationships between constrained parameters. There is significant merit in fostering a climate of curiosity and creativity across the overall enterprise, allowing individuals more room for independence in the big data and decision-making space.

Equ : 2 Real-Time Inventory Level Update Equation

$$I_t = I_{t-1} + R_t - S_t$$

- I_t: inventory level at time t
- R_t: received shipments at time t
- S_t: units sold (or demand fulfilled) at time t

4. Technologies for Big Data Engineering

An exploration of big data practices in the retail sector has been conducted by. All of them emphasized solely the technology side and mainly limited their focus on the applications and consequences including responsibilities. Consequently, a gap in the literature existed on grasping the phenomenon in-depth. To fill the gap, the real practices of organizations in the retailing sector is discussed. Central issues related to data sources, collection process, processing, analysis, and its integration in operations are illustrated. The role of technological development in creating this phenomenon is highlighted. The technology is available but needs to be applied. If the process is not integrated, people may interpret data for their own interest and arrive at different results. Data from point-of-sale was viewed as the most valuable by all firms' practitioners. The data was collected from long before, but mostly processed as snapshot data with no personalization or weekly range, and further analysis was almost non-existent.

Closest to the ideal situation, actual and sensor data were intertwined by the fastest firm. Since data from sensing devices needed to be converted to a readable format, processing was mostly software supported but limited with simple metrics. Data was processed during reporting periods but analyzed over a longer period of time due to the necessitating complexity. Results are acted upon immediately, but needs further analysis for improvement. Several levels existed in terms of desired and existing data types but all targeted the instant and seven-meter or even real-time data. Most of the firms possessed vast data, but data from previous years or databases were mostly planned to be applied. Efforts were on-going to implement transaction-level data mining tools, but conflicting views existed on the information desirable from the collected data.

Timeliness refers to both the data processing and the physical capability. The most wanted actions included price relabeling, assortment change, sales low-level tracking, staff efficiency, and improvements. Decisions around availability, assortment, pricing, and layout planning are identified as key retail operations that can benefit from more advanced data processing and analysis. For usability purposes, a wide range of applications including centralized control, perception testing, instant custom preparation, and communication were envisaged, but few had launched or planned implementations. Existing applications of big data in retail operations concern sophisticated analyses and access to high volumes of data in a short period of time. Few applications existed for operations decision making. Some applications are either still conceptual or planned.

4.1. Data Storage Solutions

In this research, the data had to be uploaded in different physical locations so that newly formed UPH data and sales orders could be registered as fast as possible. This meant that data was constantly uploaded within a few minutes interval for each WH. The fact that data from the AS had to be uploaded to multiple servers also increased the physical data transfer between the servers. There were two main scenes of the processing units. One consisted of filters that aggregated data on a WH basis, which allowed all the models to be executed in advance on the previous day's data. The other processing unit only executed the last two hours' data on the same day to form DSO and part of the QP that would be sent to UPS before 0930. This was suggested by students due to time constraints. This movement helped to ensure the availability criteria when otherwise data had to be uploaded in less than seconds, which, based on preprocessing and arbitrary data sizes that would be employed, was semi-impossible.

For better flexibility, data were stored not only for optimal model parameter tuning or result evaluation but also the raw historical data were kept to help evolutionary algorithm evolution trend track and outlier definition progress. Due to this constraint, the data storage of the data behind the visualization also had to be changed. Per multiple WHs, data were accumulated in a finished dataset and filled based on a rounding-up time that was divided by both 5 and 8, just like the raw data. So even for 90 days data monitoring, only 90 * 48 (5 min approximation for 8 WH) finished datasets would be created and stored. Due to quantity constraints, finished datasets were saved in txt format, each containing data for only one WHO one month back. This allows the aliasing of servers in the BSD, where at least one of the servers would always be able to access the finished dataset and simply filtering it based on emptying makes the access time less than a half of seconds.

Besides this, to help with debugging or checking model convergence issues, raw historical data, processed daily summary data, and interpretation result set data were also stored. Instead of keeping these datasets still in txt format, these datasets were all uploaded to the BDS with commonly used Parquet format, where data files were sorted based on the daily used rounding up time. The specific kept datasets will be presented in further sections when explaining the whole visualization pipeline in detail.

4.2. Data Processing Frameworks

Real-time inventory optimization in wholesaling requires monitoring of large, varied data sources to continually renegotiate sales commitments and communicate changing expected delivery dates to multiple customers. Big data technologies have emerged to meet the scale requirements of data processing tasks, addressing computational and operational challenges faced when applying traditional data processing tools. BDE requires data processing frameworks that are distributed to meet scalability and reliability requirements and support fault tolerance mechanisms and connectors for cloud environments. Four frameworks are evaluated for suitability to wholesale BDE use cases: Apache Storm, Apache Spark, Flink, and Kafka Streams. Stream processing engines process data continually on an unbounded data stream or in bounded data batches, and relevant frameworks fall within the distributed stream processing engine category.

Wanting to focus on a single cluster deployment on a cloud environment, Apache Spark and Apache Flink are eliminated based on platform complexity. Cloud-managed services for Apache Kafka and Apache Pulsar act as full-fledged offerings, providing a controlled experience expected to achieve better performance. Flink's adapter for Apache Pulsar is found multi-layered and unnecessarily complex for onboarding/existing clients already familiar with Pulsar. Spout implementations for Apache Pulsar within the Storm ecosystem are found unreliable, requiring custom-built solutions likely lacking extensive peer review. Consequently, Pulsar is rejected in favor of Kafka Streams. A data processing pipeline targets Kafka Streams through a combination of data connectors and the Confluent Platform implementer. This pipeline is designed generically to be vendor-agnostic and modular in enabling supplemental integrations. Connections to event sourcing storage mechanisms supporting eventual consistency and adapters for communicating with orchestration engines allow for flexibility in sourcing tasks.

4.3. Real-Time Data Streaming Technologies

It is essential to choose the appropriate technology components that fully satisfy the application requirements. To address the real-time streaming context the proposed solution includes data decoupling, data access and retrieval, data preparation, data storage, stream processing engines, batch processing engines, and data presentation layers. These technology components are scanned and suitable ones are chosen for implementing the proposed system architecture. Here the following technology components are analyzed and chosen the ones that are applicable.

4.3.1. Data Streaming Sources: Data access and retrieval technology choices comprise the data streaming sources such as data decoupling, data access and retrieval, data preparation, data storage, stream processing engines, batch processing engines, and data presentation layers. Here Kafka is chosen as a message broker for decoupling producers, message subscribers, and streaming processing pipelines. For real-time data streaming, Kafka is widely used. It increases scalability and provides simple APIs for developing message producers. Kafka also has strong Good-to-Know capabilities. But a delivery guaranteed query implementation may need some effort to achieve the Exactly Once semantic applications.

4.3.2. Distributed Streaming Data Processing Engines: The selected data processing engines for real-time data processing streams should satisfy the characteristics as batch streams and micro-batches. In the batch process focused solutions Spark and Flink are both spelling and good solutions. Spark is supposed to process the streams in micro-batches that delay the latency to a few seconds. The delayed presentation of data in a macro stream can be avoided with Flink. Compared to Spark, whether batch streaming processing is performed, the selected engine would be Flink. For fast-moving streaming data systems, it is crucial to maintain the data streams at a very large volume. Some technologies such as HBase, Cassandra, and Big Table were vertically scaled at this level. Blob storage technology applies to horizontally partitioning and scaling the system. But at this level, OLAP Data Warehouse technologies provide fast reads with very large volume data. And OLAP engines allow implementing the analytical systems. Since the implementation of the proposed system works at a different system speed, it is not suitable for benchmarking and testing the research results.

5. Data Collection and Integration

In essence, big data refers to large, complex, and fast-moving datasets that cannot be analyzed using traditional techniques. Recognizing big data marks a turning point in industry competition. The perceptions of supply chains have been improved by big data development, establishing speedier responses, better collaboration, more customized supply chains and greater inventory turnover in the supply chain, etc.. Nowadays, revolutionary changes are brought to existing systems by all kinds of internet services. With the widespread advent of various sensors and devices, a large amount of data is collected and utilized to optimize and improve efficiency of the supply chain. However, similar to traditional data, these big data sets still need application, especially running analytical tasks to provide information or insight for the decision making processes.

Using a multi-dimensional big data collection and integration platform to optimize inventory in the supply chain is the focus of this research. It involves the decision-making perspectives of how to virtually and flexibly manage inventory in a distribution network. A framework to combine NoSQL technology, analytic computing and visualization services to acquire and analyze heterogeneous data is proposed. The compatibility of system components with the IoT architecture to set up a big data environment is also examined. In response to these challenges, an innovative methodology to design big data collection and integration systems by adopting hybrid NoSQL databases, analytic computation and visualization technologies is developed. To support the functionality of the proposed system, multi-dimensional architecture, software and hardware requirements, data models, as well as data collection, integration, storage and processing schemes are discussed in detail.

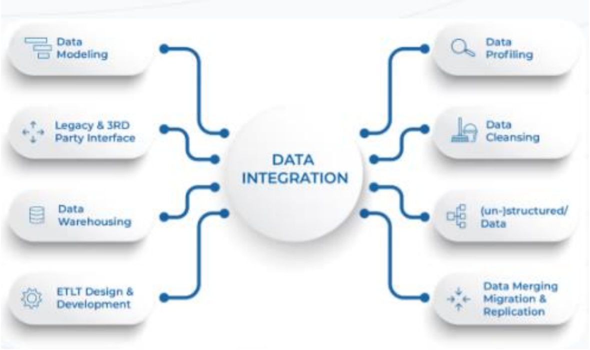


Fig 4: Data collection and integration process

In addition, a case study of running design and analytical applications of the platform to provide insights into inventory optimization for a product in a distribution network is demonstrated. With the multi-dimensional big data collection and integration platform, intelligent business is obtainable by examining different analytics to provide anti-stockout insights in distribution networks under mixed distribution strategies with 1-hour processing time. Having promoted the event study analysis of market demand, warehouse inventory, transportation routes, and price range of the studied product, the level of systemic intelligence in making real time and long-term decisions for supply chain operations will be enhanced.

5.1. Data Sources in Wholesale Distribution

In wholesale distribution networks, various data sources are exploited to characterize the state of the business with regard to inventory optimization. The primary data sources including the business's product catalog, supplier and customer databases, demand forecasts, customer orders, inventory and battery stock records, and expert knowledge regarding policy and constraints are elaborated upon in what follows.

Wholesale Distribution: A wholesale distributor collaborates between suppliers and clients. In order to satisfy customer demand, a distributor must dispose of an adequate stock level of specific products. To minimize costs, the inventory optimization problem must set inventory levels while taking into account stock holding policies, service levels, and supplier lead times versus costs. Inventory Policy: For each product of catalog, an inventory policy defines both stock-holding levels and stock-procurement actions. Each policy must comply with the constraints set by the service-level norm, the authorities' regulations, and the supplier agreements. Policies for bulk liquor and other selected products allow only for a two-way change of stock holdings. Distributor Product Catalog: It comprises information about the product states and resource definitions regarding associated data structures and record entries.

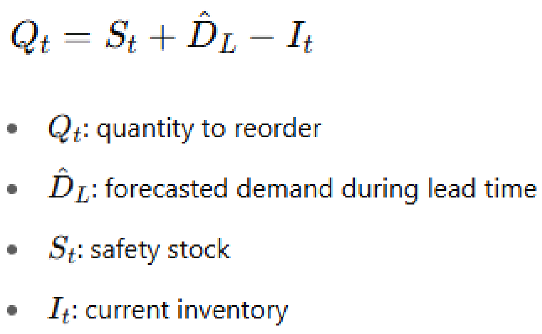
Supplier Catalog: This source stores details on the suppliers such as location, stock-holding pricing, advance communication requirements, blocking dates and capacities, portfolio states, exclusivity agreements, etc. The normalization of supplier records to include limits in points allows avoiding data redundancy, relying on supply portfolio states instead. The effective algorithm is then used to maintain weights of this relation in order to make supplier classifiers and clustering models. In addition, the queries deemed to be impractically long are substituted with a collection of simpler return queries. Batches of orders are determined by each product category, product, for unit orders only. Stockholding batches are determined by resource and current level ranges.

5.2. Integration Techniques for Real-Time Data

Data integration is a significant problem in big data engineering. Data integration is classified as a schema-based integration and non-schema based integration problem. Schema-based integration of inductive, record, relational schemas into XML is proposed. Message based architecture is classified into sending nodes, receiving nodes, messages, and recovery mechanisms. A model is designed for message writing with needed support in 4 methods. Publishing and subscription formats for traditional messages and publishable stream processing in various window-based time intervals is designed and defined. Using log file source, a publishable stream publishes event observations per 3 minutes. The P1 and P2 pub-sub subscription formats are used. For real-time processing or integration of data streams in ETL, a real-time ETL with appropriate techniques is proposed for continuous data sources. A visualization tool is designed that performs text mining to identify the theme of journals and categorize them. Data streaming is primarily data integration and considered as a dimension-less input of a warehouse for processing.

Data integration is an important research area in ETL and a semantic web marriage of the ontology-based integration schema and design is experimented. Though the proposed architectures deal variously with data model, schema, and information integration, a focused loaded data and ontology mapping integration architecture is designed addressing the limitations of existing architecture. An ETL framework with data streaming concept is defined and various levels of data streaming sources are categorized with an extraction approach relevant to the levels. Distributed ETL environment data streaming source compatibility is defined and ETL models are developed for real-time data warehouse compatible with SCHEMA-BASED DATA INTEGRATION ARCHITECTURE. A conversion function with concerned design steps for converting the extracted queries to loadable ones is designed and implemented. A new schema which is the observation of the description of all pre-accumulated matching sources, already defined schemas or derived independently is discussed. An initial data selection procedure is proposed, which satisfies this schema-based matching solution without examining the documents or data with a space complexity.

Equ : 3 Inventory Replenishment Decision (Order-Up-To Policy)



$$Q_t = S_t + \hat{D}_L - I_t$$

- Q_t: quantity to reorder
- $\hat{D}_L$: forecasted demand during lead time
- S_t: safety stock
- I_t: current inventory

6. Data Analysis Techniques

As the sales forecasting engine remains dependent on historical data along with statistical and judgmental techniques, predicting the level of inventory needed to avoid 'out-of-stock' or 'overstock' situations is difficult for replenishment managers for a store on and after a snow event. Colder or rainier forecasts likely impact on shelf-stable dairy sales while competing will not. and this is doubly perplexing, as snow would tend to push demand for certain items lower, while at the same time push demand for other items higher. For a wholesaler, daily sales forecasts are required at all products sold at different retailers, as regular procedures become ineffective due to modeling errors and processing delays of existing data transformation techniques as well as entrenched demeanor of price prediction on ill-conceived homogeneous movements. The dependency on historical data and the associated difficulties could be addressed by exploring the additional data sources and identifying the relevant factors impacting sales. This involves resolving the problems such as non-availability of factors due to confidentiality policies of very large retailers and/or data extraction difficulties from quick and dead rows of available public data. The next big data challenge faced is that of investigation speed of prediction as transformation of huge amounts of data sets stored in numerous lines across many tables. The solution for prediction accuracy is to adopt relatively complex deep learning methods which require substantial hindsight to train on a considerable amount of data.

A considerable understanding of the underlying sales phenomenon is also required to derive any useful input factors. However, all of them would be ascertainable, while definitive insight preparation and interpretation of always non-straightforward prediction results would remain big data challenges at wholesaler side leading to exorbitant costs, profound enigma and much frustration. Hence, wholesale product level sales were re-casted for accuracy by adopting partially accessible online and semi-public social factors including weather and internet search trends along with goodwill proxies and direct price data of both wholesaler and retailer. And big data analysis was successfully discharged via methods originating from and hired by the business intelligence domain. The high hierarchies of grouped variables were obtained via sensitivity analysis on a designed unsupervised market tree/model as the relevant input of multi-dependant concurrent variables for a gross accuracy improvement. The investigation domain for prediction could be temporarily reduced into partitions of key products and retailers to expedite prediction speed. At least linear correlation and significant variables candidates for prediction of sales were computed and discarded leaving thousands of product-store daily time series sales for model training to obtain hourly prediction via optimized deep recurrent neural networks. During the stage of prediction re-casting, instant direct price change prediction of each retailer would be available by its archived data transformation methods.

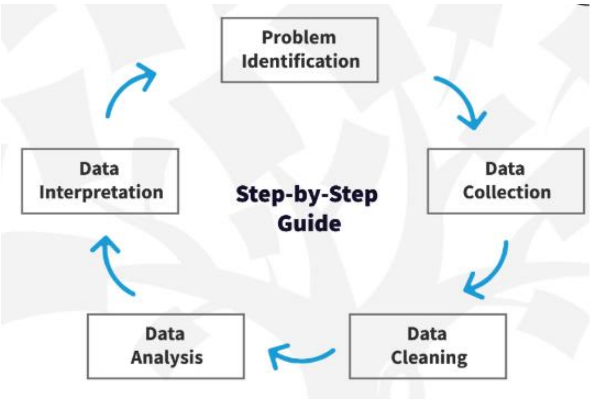


Fig 5: The Data Analysis Process Step-by-Step Guide

6.1. Descriptive Analytics

Assessing the state of the network is one of the first steps in the inventory optimization process. In this step, not only qualitative but also quantitative data of the network will be thoroughly scrutinized. Using the facts gathered in the previous phase, the wholeness of the supply network is observed from the aspect of distribution and optimization. Most of these factors will generally remain the same; however, factors such as order volume and purchase price might exhibit different characteristics at different times due to changes in the wholesale supply market caused by environmental changes. The collected factors will directly impact the accuracy of predictions and therefore the quality of inventory optimization . Hence, this phase will require continuous observation of the factors involved in order to provide a clear insight into analyzing and estimating the system.

A quantitative overview of the daily incoming orders (DIO) of the city will be required to realize the gradual volume of orders of the product groups. As the main purpose of real-time inventory optimization (RIO), it will be essential to view the number of orders and key performance indicators (KPIs) such as average order volume or behavior of the landed purchase price trends over the week. Most of the states will be viewed at a weekly level to analyze trends and volume differences at an aggregated level, which is useful to observe but not straightforward or tractable to analyze.

To observe the general trend of distribution from a wider perspective, the change in the number of incoming orders (DIO) has been visualized with its change over time calculated daily. The volume change in an individual city can be seen widely, along with its percentage change, which can be particularly useful in providing a general understanding of the order fluctuation due to behavioral changes in the market and how adverse the change can affect the network.

6.2. Predictive Analytics

Demand prediction deals with estimating the future demand for a product. Accurate demand prediction reduces a retailer’s stock costs, such as over-stocking, under-stocking, and restocking costs . Quicker adjustments to a retailer’s future demand increase forecasting accuracy. In retail distribution, demand prediction can be applied at various periods, such as hourly, daily, weekly, and monthly. Demand prediction models apply an extensive amount of predefined factors and exogenous events. However, overnight demand prediction requires more forecast preparation time with these models. Thus, exogenous variables, aggregations, and dimensions needing accurate demand prediction should be defined clearly before addressing the problem. The demand prediction preparation process can be solved with an automated process to extract and check factors and checks for pre-aggregation.

Many existing models try to solve this situation accurately and promptly; however, such exogenous variables, model construction preparation, defining periods, dimensions, and categories are not defined clearly in these research models. Models like a filter or variable selection methods that can help construct demand predictors with less effort would be useful to researchers and practitioners. In addition, no model with a feature selection method is proposed along the whole demand prediction construction process. Therefore, such automatic model construction with feasibility checks as the preparation process of a wide variety of models is valuable to researchers in demand prediction, plus it can be used to extract new markets and products easily. Many existing models choose a few candidate factors for a product and use conventional models to check and fit these factors. Still, they do not classify importance and check factors before candidate factor selection approaches.

6.3. Prescriptive Analytics

Prescriptive analytics uses big data, statistical modeling, and optimization techniques to recommend the best course of action to improve performance. Linear programming and constraint programming have been used for many years to optimize various applications like LOT/LIN scheduling and inventory management. The advances in data science and big data have spawned several big data-enabled prescriptive analytics innovations for various applications. Due to powerful and low-cost computing infrastructure and data storage, modeling and data analysis of unprecedented scale have been motivated. On the modeling side, new optimization techniques have been developed to better model very large-scale data. On the data side, many new AI techniques have been developed to learn from big data effectively.

A prescriptive analytics framework to combine both descriptive and prescriptive analytics. New innovations in data science and optimization are introduced to turn descriptive analytics products into end-to-end prescriptive analytics solutions. A new machine learning approach to tackle the big data problem in understanding the nation-wide retail environment is presented. There are three important steps. First, point-of-sale data from every single store in the nation is used to build a dynamic Bayesian network (DBN) model for multivariate time series prediction. Second, predictions from the model are fed into an integer programming-based optimization model to deliver optimal promotion execution timing. Finally, an innovative big data visualization tool is built using the state-of-the-art D3.

Innovation in real-time big data analytics. It is not necessary to provide prescriptive analytics on optimal purchase targets. Instead, intelligent information on violations of optimal targets can be disclosed. Machine learning techniques are used to study aspect sensitivity in intelligent pricing. Recent advances in natural language processing (NLP) technology are used to identify the pros and cons of investments in a large number of news articles. Based on big data and data-mining discoveries, hospitals are segmented to assist inventory and pricing decision making. A simple yet effective approach is provided to enhance consumer engagement in applications where consumer price elasticity is sensitive with respect to recent pricing history. Applying the analysis framework yields better performance than in-house proprietary technologies.

7. Machine Learning in Inventory Optimization

While traditional inventory optimization techniques have proven effective in various scenarios, they often fail in complex supply chains with many items, locations, and constraints. Consequently, Machine Learning (ML) and Artificial Intelligence (AI) methods are now being actively investigated to enhance inventory management. ML techniques can be categorized into supervised learning and unsupervised learning based on the nature of the training data and prior information available about the problem. Supervised learning requires historical input-output pairs for training, and this category includes methods such as regression, tree-based approaches, recurrent neural networks, and deep learning approaches. On the other hand, unsupervised learning explores the underlying structure of the world and derives approximations of data-generating processes without need of any output.

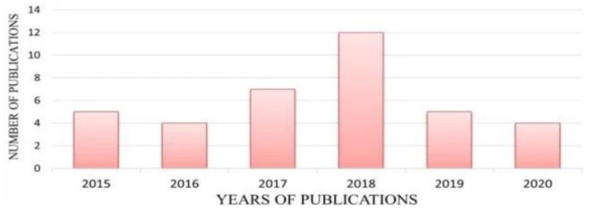


Fig 6: Big data engineering for real-time inventory optimization in wholesale distribution networks

ML has recently gained popularity in various domains, including supply chain management, owing to its impressive performance in accurately predicting Inventory Holding Costs (IHC) and summarizing demand. Moreover, its performance has been shown to improve continuously with additional data. The applicability of traditional machine-learning methods, such as regression and gradient-boosting trees, in estimating IHC, is examined in addition to addressing ensemble-learning methods, such as those based on random forests, which can take both numeric and categorical inputs. Moreover, the application domain of energy storage analysis and the need for efficient, scalable, and interpretable methods due to the underlying data's size—single-agent, single-product, and continuous meta-observation horizon—are discussed. A comprehensive literature survey of decentralized inventory control in supply chains is conducted by reviewing and classifying 314 research papers across relevant fields based on domains, models, techniques, and future work.

7.1. Supervised Learning Models

Univariate Logarithmic Regression (UGLR): UGLR predicts the logs of the time series as a linear function of its own lags. The mathematical definition is defined as \[Y_{t} = \hat{\alpha}_{0} + \sum_{i=1}^{n_{p}} \hat{\alpha_{i}}Y_{t-i} + \epsilon_{t}\] and volatilities can be predicted by the standard deviation of this regressed residuals of \[\sigma_{t}^{2} = \hat{d_{0}} + \sum_{j=1}^{n_{q}} \hat{d_{j}}\epsilon_{t-j}^{2}\]. This method needs more than 10,000 collective state (batches) data to train the parameters.

Multivariate Granger regression (GRG): Let \(K\) be numbers of observation time steps. In this study, it defines \(y_{i} (1 \leq i \leq n)\) as the variables of each inventory groups, \(t\) be the time index, and \(L\) be the order length of regressors when \(n_{o}\) is larger than setting upper bound and \(d\) is the truncation delay when \(n_{o}\) is smaller than it. The mathematical expression was defined like this as equations \[\sum_{j, j \ne i}^{n} p_{ij}(y_i|y_{j,L},...,y_{j,1}) = o\] with \[\sum_{j=0}^{L} \sum_{k=1}^{d} \sum_{n=1}^{d} \phi_{jk,kn}x_{j,k}(t-n)y_{i,L}(t-l)+\epsilon\] which implies that the affected inventories have multiple closure conditions while they are observed at the same time if they are sufficiently closer to keep inventory. VAR was first proposed in a strict form in and similar formulation was also used to resolve anomalous time series in another form which does not need training data directly. This method can take structural parameters from time series data directly without knowing the precise stochastic processes.

Unconditional average: A simple approach for forecasting the future value is to use the historical average as the approximate. In this method, it is assumed that the strength of the relationships between the variables remains unchanged comparatively during the forecasting span. Therefore, just the average is used when estimating the post-observed data.

7.2. Unsupervised Learning Techniques

Unsupervised learning provides a diverse toolkit useful for combing through data looking for new insights or patterns. One popular approach is data clustering or grouping similar records or events based on statistical properties. Clustering tries to find similarity in data records using models like K-means, Gaussian mixture models, or some variations of these. A clustering algorithm can build a model to group records into several clusters and flag any new records that do not fit in the distribution of the clusters as well as outliers or anomalies. Outlier detection can take an even tighter approach into only looking for records that differ from the "normal" group significantly enough to raise concern. Inner workings of matching methods also tend to be obscured from view, as elaborate series of matrices go through different transformations, but can readily provide several thousand keywords or features and a raw analysis of which records match conditions.

Dimensionality reduction compresses data by filtering out unneeded features or records using methods such as linear discriminant analysis, Laplacian LDA, and t-distributed stochastic neighbor embedding. It should be noted that this is a lossy compression stage that should minimize distortion between the higher and lower dimensional representations of the data. After the dimensionality reduction, it is common to provide intermediate charts or pictures of the results or to use higher level descriptive statistics to summarize the trends or remaining features. The reduction step should also be cost effective in memory and runtime constraints, as higher dimensionality may flow into further analysis and exponentially slow down run conditions. K-means clustering or other similarity functions can also be run after the dimensionality reduction.

7.3. Reinforcement Learning Applications

Reinforcement learning (RL) has been used for various inventory management problems in recent years. This section briefly reviews applications of RL algorithms in literature related to inventory.

IAgannoccaro and Pontrandolfo propose a general formulation for the inventory management problem with a non-Euclidean metric space. The state-space is discretized into finite states and a Monte Carlo tile coding approach is used to represent the value function. The author considers a Markovian nature of such inventory system dynamics and formulates the problem as a Markov Decision Process (MDP). A modified version of the Q-Learning algorithm using the linear value function approximation is proposed to solve MDP. The proposed approach is tested in simulation in a non-stationary environment characterized by time-varying demand and lead times.

Case-based reinforcement learning (CBRL), a hybrid learning technique, in which case-based reasoning (CBR) and RL are integrated and work together, is proposed to improve decision-making in uncertain and dynamic inventory systems. The dynamic inventory control problem in a multi-agent environment is used as a benchmark problem. Simulation experimental results demonstrate that CBRL can acquire policy networks with considerable quality in a short time, though the quality is less than that of the Q-learning policy network. Moreover, further training can effectively enhance the performance of the policy networks.

Another approach is proposed to integrate genetic algorithms (GAs) with RL methods to improve the ease of implementation and interpretability of policies.

8. Case Studies

Network data is the cornerstone of the modern business environment and, as a result, it is predicted to grow at an unprecedented rate. Effective network data management is a key concern for business professionals. This paper focuses on the wholesale distribution supply chain, which has unique network structure and data types. The challenges and methods to effectively obtain and analyze data are discussed using business data as an example. Specifically, real-time product information, wholesale distributor information, retail customer information, sales information, inventory information, and granular data on sales prices and costs were obtained.

Data-analytic techniques such as real-time demand forecasting and inventory optimization can be applied in the wholesale distribution supply chain network to leverage the value of data and enhance real-time operational excellence. Specifically, local inventory optimization (LIOpt) can be derived for stationary systems with periodic inventory replenishments. Other advanced analytic techniques such as static inventory optimization are applicable in enhancing the value of data and assisting managerial operational control decisions in the wholesale distribution industry.

Customers of the wholesale distribution supply chain network have multiple-tone demand. They regularly replenish their inventories by ordering products from wholesale distributors. As product lead time is large and important, pickup trucks of wholesale distributors periodically deliver products to unclamped bins of customers, and customers receive products at dock doors. The average demand distribution pattern per customer can be identified for demand forecasting for each pickup week (5 days) and inventory optimization for each operation day. Intuitively, local cost/minimum inventory is incurred when supply equals demand, and local LIFO pricing from customers induces 1 PPIC local inventory optimization.

8.1. Successful Implementations

During the last few years, we have witnessed several successful implementations of multi-echelon inventory optimization software in Europe and North America. In the first implementation in 2008, software was developed to support near real-time inventory pre-positioning optimizations for a leading global fast-moving-consumer-goods company. It involved a series of implementations moving from validating the amount of optimization that could be allocated to a new system via stock holding positioning software and visual optimization dashboard linked to a new data warehouse for a European subdivision.

Within a few years, the project transformed to encompass the slightly larger inventory category or prepositioned stock locations, the relocation of inventory across the warehouse network, and co-selling of two brands, all blending various technologies with a geolocation view. Following a demotivating long re-implementation with a fresh team over a year, the project finished in 2020, with a new priority assigned to the next decade's wholesale vendor management inventory prediction and trust-based supply chain replenishment systems, which involved the consideration of handling the competition from pure e-care. While this team will touch on many of the new algorithms learned, all work is based on cloud systems with implicit rewards, implicitly complex inventories and distribution networks, and on optimizations made less than a day before receiving a new updated forecast.

Another notable implementation is stock allocation in a lump selection product distribution system for a leading European centered price fashion retailer. After a multi-competing system-wide convergence to a promising statistical prediction approach over nearly a decade, a fresh project was initiated to blend prediction with allocation over a two-day window with possible multi-location announcement in collaboration with the algorithm developers for individual product items at the central capacities, on top of the success of a standalone decision support system. Though substantive enhancements were made to prediction and allocation accuracy and granularity to include shipment batching and price adjustments, real-time uses against a database housed build expectancy tables have yet to be available.

8.2. Lessons Learned from Failures

This dissertation investigates multi-echelon and multi-stage inventory optimization in general distribution networks in the presence of stochastic demand, lead-time, and yield. Based on a discretization method, the hybrid solving procedure combining Lagrangian relaxation and branch-and-bound algorithms is developed for a trusted network. Furthermore, heuristic control mechanisms in the form of threshold policies are proposed for the distribution network operated by retailers, which is subject to demand-side load fluctuations. To tackle the problems in the cloud-based inventory management systems with real-time workflow, a parallel batch dynamic programming algorithm has been developed to deliver the optimal stocking strategy in the shopping cart pipeline.

The major achievements of this research have been highlighted, along with its innovative framing of the problem. The approximation and buffering flexibility techniques pave the way for a tractable small-scaling algorithm. The resulting two hierarchies of distributed optimization methods show their great potential for large-scaled networks, especially in the era of cloud computing. Future research directions have also been proposed, targeting the problem instances involving dynamic edge-weighted graphs, as-limited-as-necessary human cognition, and so on.

Many lessons were learned during the project. The first key lesson was that small experiments with proper metrics should be run before embarking on a big project. Early experiments are very important within the project as a way to highlight failures and reload the project or make necessary changes. Without this, the project may go on indefinitely then run into major issues and fail as a whole. As an example here, a limited test dataset of randomly selected SKUs from one small region was created to evaluate the importance of point of sale data and time series embeddings concerning network design.

These early analyses provided sufficient insights into changes to be made during implementation.

9. Challenges in Implementing Big Data Solutions

While wholesale distribution companies seek the business value of big data technologies, they encounter great difficulties in implementing these big data solutions. An object-oriented design-based global data model for integrating and managing data from disparate sources adopted in order to administer a batch processing data lake architecture. The design process, technological details, and a demonstrative example regarding integration and management of heterogeneous pricing data in a big data environment have been illustrated based on the proposed data model. A feasible, reliable, and expandable solution has been provided. Its impact on quality and performance improvement of the system can be acknowledged after being discussed in detail.

However, there are some challenges in implementing this solution. To begin with, the usage of this new system needs great adjustments in the task performance of the data team. Since all the tables were totally stored in spreadsheets with long column headers in previous practices, the transition to the new logic structure of data assets necessitates varying operational and processing techniques for different data sources or systems. For example, only with table headers: [‘base_pgid’] or [‘quently’] in the table from a simple data source, it is impossible to identify duplicate entries of distribution ID (in terms of [‘base_pgid’]) or observability of forecasts (in terms of [‘quently’]). Team members also require appropriate training to use the data management tools effectively.

Additionally, prior to the integration process, a large amount of outdated, irrelevant, or interruptive data in the existing database should be removed. Also, numerous tables without identification of data sources or inception dates exist in the current database, which hinder an accurate integration process later. Furthermore, missed or incorrect values in the raw data need great efforts to harmonize. For instance, regarding wholesale locations, various format types and filter characters are all recorded in the same cell.

9.1. Data Quality Issues

Big Data provides a wide range of possibilities and challenges in utilizing data within wholesale distribution networks. The importance of the development of advanced data analytics and technical knowledge is needed to gain insight and understanding from the data. Such insight can then be leveraged to enhance supply chain decision-making capabilities, including better forecasting of demand predictions, inventory optimization processes, infrastructure utilization, and improvements to the distribution network structure planning. However, there can be challenges in regulation, governance of intellectual property rights, and technology adoption-related issues, which can impact the quality of the data. For the sake of a thorough understanding of the prediction and optimization processes, the data pipeline with consideration of data visualization is also important.

Although this thesis is focused on mathematical modeling and algorithm development, it has to start with the data representation within inventory optimization. Challenges and opportunities related to data quality need to be addressed and explored–data quality can directly affect the performance of the process and the prediction results. Institutions, either in academia or enterprises, all have their own information systems that have evolved organically over the years. This might lead to differences in software or data representation solutions adopted. The effects of these multi-systems on machine learning have not been fully understood yet, whereas the rapid evolution of information systems and the fast growth of applications that involve the usage of them push this to be a key element for the success of data-related research in not only enterprise environments but also the society at large.

Regarding data consumption and machine learning, data quality issues from different perspectives are reviewed and illustrated with the example of wholesale distribution networks. After that, four aspects–data inconsistency, missing values, outliers, and textual data–are discussed in more detail, along with the possibility and need to mitigate the issues in wholesale distribution networks. However, as complicated application systems can incur difficult-to-tackle data quality issues, the insights introduced here serve as a framework reference at the level of theory and methodology. It is hoped that the detailed approaches and solutions can be provided, in either state-of-the-art or in specific variations, in future collaboration between academia and industry.

9.2. Scalability Concerns

Real-time inventory optimization for bulk stocks and perishable product is extremely hard. The system might have hundreds of locations. Further, each location has hundreds of SKUs each with multiple product characteristics which complicate the system even more. On top of that, with continuous arrival of big data, and forecasted/out of stock input pushing through the system frequently, real-time optimization of safety stock levels and daily replenishment schedule is extremely hard. In this study, a scalable two-layer optimization framework is proposed addressing safety stock allocation and daily replenishment scheduling optimization separately. To cope with real-time transparent big data of a distribution system, newly designed data pipelines for cleaning and storing the input data, as well as calculation of 3D inputs are proposed. They transform data reservoirs into an insightful big data warehouse feeding the big data engine for inventory optimization. After 2D safety stock calculation on the SC level, improved daily replenishment plans are redistributed to branches to operate best with other inventory plans an hour ahead of time. In the large-scale case of a real-world distribution system, the two-layer optimization framework provides reasonable and unexpected safety stock allocation and delivery plans with good improvement on inventory performance. The results are also verified with Monte Carlo simulation across real-world scenarios. Meanwhile, the safety stock optimization and daily replenishment scheduling are evaluated against continuous stochastic demands or other uncertainties. Enhanced simulation setups provide robustness testing for the inventory system. Essentially, safety stock optimization presents an acceptable upper bound on performance against forecasted demand. Safety stock allocation and daily replenishment scheduling are increasable and complementary for global benefit where mixture planning is commonly adopted in real distribution systems. In inventory management, judging whether the performance of abundance ordering policy could be enhanced by collaboration or what is an optimal collaborated delivery mechanism are more interesting. There are abundant studies addressing such questions by simplifying either demand or supply sides behavior. However, the key question remains unresolved, examining ordering behavior under Nash and fair-shared delivery insight when all agents have a fair share of consideration. It is hard to know, as they imply a GP-type solution that is famous for its high complexity despite efficiency. The complexity could however be bounded on the biggest non divisible single delivery load.

9.3. Real-Time Processing Limitations

The wholesale distribution automation (WDA) process can be modeled as a data pipeline that consists of six primary subprocesses: (1) order allocation, (2) incoming merchandise receipt, (3) work queue creation, (4) work execution, (5) work confirmation, and (6) work completion. Each subprocess can be implemented and performed by various logical and/or physical systems that represent the WDA process and, most importantly, produce intermediary or final results over time as both batch views and streaming views. Such a process can be viewed as a workflow consisting of different connected operators that may trigger processing advancement from one operator/entity to another. For instance, the outcome of the order allocation subprocess can be treated as a batch view of the order allocation workflow. However, when the execution of this subprocess takes a long time, this batch view may be considered stale. Although not representing the full processing state of the workflow, the request of a merchandise drop-off from the receiving or work execution subprocesses can also be treated as a streaming view of the workflow. Instead of static binary values that indicate the completion or dissolution of the overall workflow, the wholesale distribution (WD) domain requires a case-dependent tolerance and processing time for viewing the staleness of its batch views. As such, it is a unique and challenging research problem in big data engineering, i.e., how to manage the viewing staleness of both the batch and streaming views of a WDA process in real time? To address this challenge comprehensively, three perspectives are taken.

First, there is a need to present a theoretical study of view management that captures consistent ground truths of streaming views of a WDA process and its case-dependent tolerance for batch views in real time. Second, a set of well-defined algorithms to determine the viewing completeness of various types of views and a detailed computational evaluation of the proposed algorithms with an illustrative example are provided. Third, issues regarding potential application scenarios in the WD domain and performance evaluation of the proposed algorithms on real data collected from a large global wholesaler are discussed to provide insights into the impact of view management on the success of any realistic applications.

10. Conclusion

In the past two decades, many technologies have been developed that have changed the way we celebrate, shop and live. Advances in processing technology have made it cheaper and easier to capture data on activities, behaviors, and transactions to support decision making. Data can be collected from login transactions, monitoring equipment, innovative kiosks, and smart meters to name a few. Advances in communication technology have resulted in the development of wireless networks, high-speed connections, and the Internet that have provided the means to rapidly transfer data around the globe from one source to another. The convergence of these two technologies has resulted in an explosion of new sources of data and an increase in the speed of data generation (starting off small and growing to become gigantic) which future research will have to deal with. Advances in computational technology have made it possible with larger storage at lower costs and more powerful algorithms to effectively analyze and use massive datasets. While these technologies have the potential to radically alter how business is done, they also bring challenges. Policy makers, business executives, managers, and academic researchers face an uncertain landscape where the possibilities are immense, but also potentially disruptive.

Contemporary literature, theory, and methodologies of enterprise/industry yield a set of questions about the implications of big data for industrial evolution. Why will ‘big data’ bestow competitive advantage? What firm strategies should be employed? Are there differences between industries and do some industries have a comparative advantage? New questions about the potential organizational, industrial, societal, and geographic implications also arise. Traditional ways of conceptualizing the organization, industry, and society become inadequate to deal with the transformative potential of big data. New economic and societal actors may emerge which exploit the opportunities unleashed by big data. New types of domination may emerge, both internal to firms and externally. Big data cannot be regarded as a panacea; they come with costs and limitations. Are (some) industries better equipped to handle big data? Are there degrees of satisfaction across industries and firms? Can the gap between the ones who exploit big data and the ones who fall behind be closed? Most traditional areas of research concerning firm performance and behavior, such as the resource-based view, strategic choice, and structure-conduct-performance, become much more complex and need to be modified accordingly. New ways of empirical analysis must also be developed to deal with new datasets.

10.1. Future Trends

The suggested future research directions for inventory systems are through a review of recent studies. Heuristic methods have been used to derive upper bounds for problems that are NP-hard in the strong sense. Future trends are developed in inventory systems, both globally and locally. In manufacturing systems, it has been shown that the variability of lead-time increases the expected downtime. Therefore, either time-based or even dynamic safety stocks should be employed. It has also been shown that temporally distributed demands tend to behave less like stationary random variables, leading to a 50% price reduction to otherwise constant demand processes. Effects of service-level constraints in deterministic lot sizing models have been analyzed in both single-stage and multi-stage systems. Moreover, a multi-echelon inventory model for manufacturing systems with a certain proportion of defective products has been developed. The combination of cost-effective and responsive inventory has been addressed. With the development of smart grids, demand-side management is a new research issue that is of significant academic interest and practical application. For instance, demand-response models and mechanisms have been studied. Following a non-linear theory, a framework based on the differential seeking particle swarm optimization algorithm has been proposed. For a fixed demand response, it has been mathematically proven that the optimal price responds to a utility maximization problem. The impacts on the day-ahead electricity market have been studied, and an analytical model for half hour-sell bonds of electricity demand units has been developed. Energy storage scheduling has been investigated, and demand enhancing methods based on the phenomenon of price changes have been studied. Further challenges have been summarized, including load change data acquisition problems.

11. References

1. Serón, M., Martín, Á., & Vélez, G. (2019). Life cycle management of automotive data functions in MEC infrastructures. *arXiv*. https://arxiv.org/abs/2303.05960
2. Pillmann, J., Wietfeld, C., Zarcula, A., Raugust, T., & Calvo Alonso, D. (2018). Novel Common Vehicle Information Model (CVIM) for Future Automotive Vehicle Big Data Marketplaces. *arXiv*. https://arxiv.org/abs/1802.09353
3. Berezovsky, A., El-khoury, J., Kacimi, O., & Loiret, F. (2018). Improving lifecycle query in integrated toolchains using linked data and MQTT-based data warehousing. *arXiv*. https://arxiv.org/abs/1803.03525
4. Capgemini. (2018). Premium OEM leverages IoT-based data in the Cloud to track vehicles in distribution. *Capgemini*. https://www.capgemini.com/news/client-stories/premium-oem-leverages-iot-based-data-in-the-cloud-to-track-vehicles-in-distribution/
5. Shiklo, B. (2017). IoT in the Automotive Industry: Concept, Benefits, and Use Cases. *ScienceSoft*. https://www.scnsoft.com/blog/iot-in-automotive-industry