ASSESSMENT 3

Acme Inc Case Study

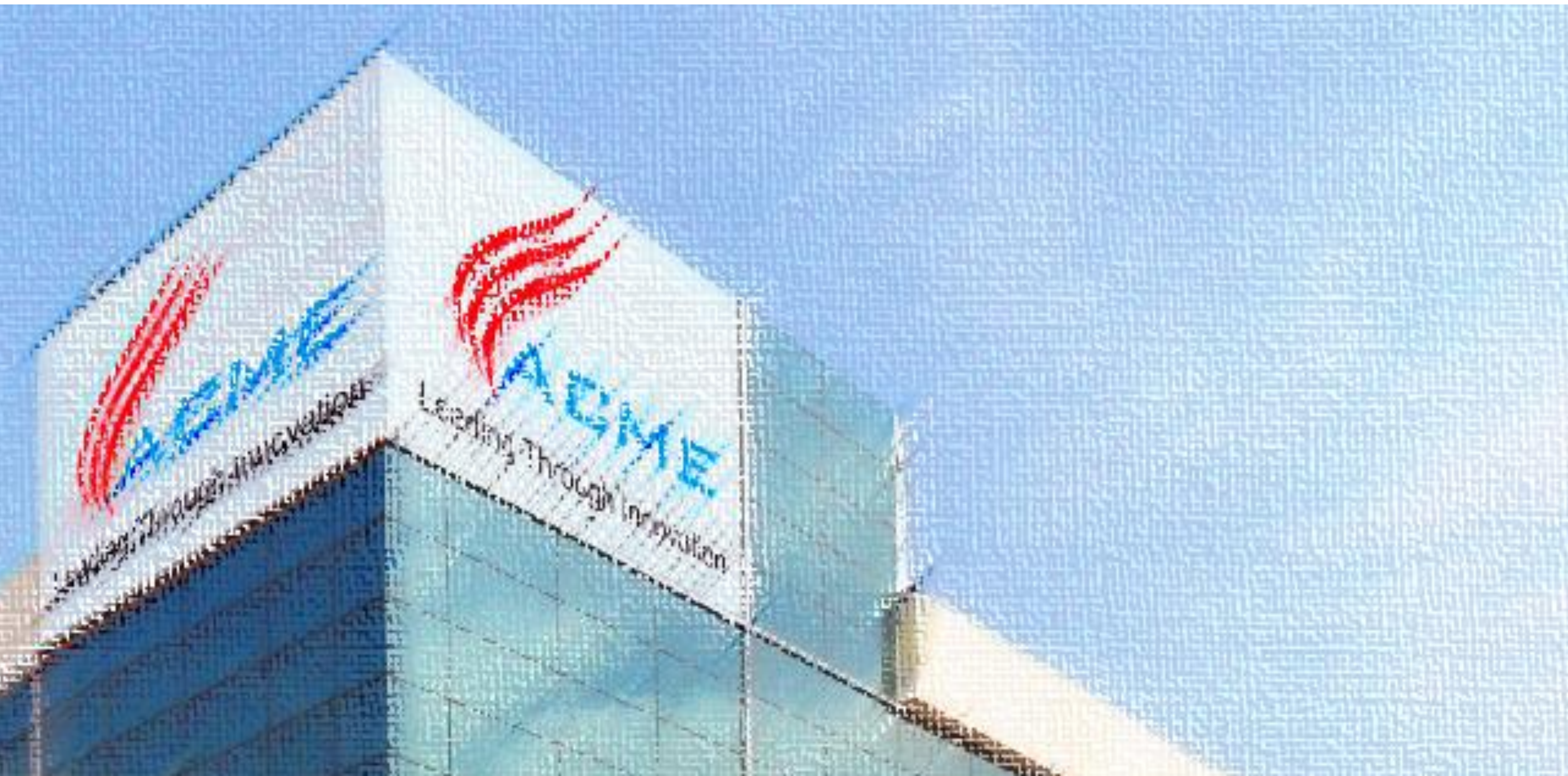


Table of Contents

Introduction 4

Current Operational Challenges: 4

Development Goals: 4

Potential pitfalls in integrating a warehouse system with existing IT systems. 5

Available Design Considerations and Which is the Best Option suits Acme Inc. 7

Sizing/Estimation and Hardware Infrastructure 8

Change Management and Communication Plan: 10

Warehouse Management 11

Improving Forecasting Accuracy 11

Advanced Forecasting Techniques: 13

Addressing Delayed Shipments 14

Architecture for Warehouse Estimation 15

Product promotion Solution 16

Data Warehouse Architecture Evaluation 19

Differences among these alternatives 20

Best benchmarks/criteria to select the appropriate and the best Data Mart/ Warehousing Architecture 22

Best Option that suits Acme Inc 22

Cloud-based Architecture pros and Cons 24

Neo4J based solution 26

Mongo-DB Based Solution 29

Machine Learning and Predictive Models for Warehouse Management at Acme Inc. 33

Conclusion 35

Reflective Insights: 36

References: 38

Table of Figures

Figure 1:Acme sizing and estimating hardware approach. 10
Figure 2:Acme Proposed supply Chain Analytics Architecture 16
Figure 3:Acme Production promotion solution. 19
Figure 4:Acme Multi-dimensional model. 19
Figure 5: Neo4j solution 26
Figure 6:ERD Diagram for Related Entities for Acme Inc. Case Study 28
Figure 7:Arbitrary(Dummy) Data on Neo4j 28
Figure 8:Acme Inc. Neo4j with Dummy Data 28
Figure 9:MongoDB-Based Framework 29
Figure 10:MongoDB Sample Database Documentation for Acme Inc.using Dummy Data 31
Figure 11: MongoDB Chart Sample Visualization (1)..32
Figure 12: MongoDB Chart Sample Visualization (2)..32
Figure 13:Summarized Steps to create ML & Predictive Models 33

Introduction

Acme Inc. is a prominent national retail chain in India, specializing in consumer electronics and durables, operating over 150 stores across major cities. With a significant market share and a commitment to providing high-value products, Acme is at a critical juncture of addressing operational challenges to maintain its competitive edge and ensure customer satisfaction (Krishnamoorthy 2014). Acme Inc. faces several operational challenges that not only impact on its day-to-day functioning but also pose significant threats to its long-term sustainability and market position. These challenges span across warehouse management, promotional strategies, and IT systems integration, each contributing to inefficiencies and impeding the company's ability to compete effectively in the dynamic consumer electronics and durables sector in India (Krishnamoorthy 2014).

Current Operational Challenges:

Acme Inc.'s operational landscape is marred by a trio of critical challenges that collectively undermine its market competitiveness and operational efficiency.

1. Warehouse Management Inefficiencies: The company grapples with forecast inaccuracies leading to either surplus inventory or stock shortages, alongside issues of delayed shipments and damage to goods, which collectively escalate operational costs and impinge on customer satisfaction (Krishnamoorthy 2014).
2. Ineffective Promotional Strategies: Despite investment in promotional endeavors, the lack of a sophisticated analytics infrastructure impedes Acme's ability to gauge the effectiveness of these efforts, curtailing the optimization of marketing strategies to boost sales and customer loyalty (Krishnamoorthy 2014).
3. Fragmented IT Systems: The heterogeneity of Acme's IT infrastructure, characterized by disparate ERP systems and databases, fosters data silos that hinder the fluid exchange of information, thereby stifling decision-making processes and operational agility (Krishnamoorthy 2014).

Development Goals:

In confronting these challenges, Acme Inc. has delineated strategic objectives focused on harnessing data-driven insights to fortify its operational and competitive stance.

1. Optimize Inventory Management: By adopting advanced analytics, Acme aims to refine its inventory forecasting, thus aligning stock levels closely with market demand to mitigate both overstocking and stock-outs.
2. Elevate Promotional Efficacy: The establishment of an integrated decision-support system is targeted to enable comprehensive sales data analysis, allowing for refined marketing strategies that resonate more effectively with consumer preferences and trends.
3. Integrate IT Infrastructure: Acme commits to unifying its IT landscape, transitioning towards a centralized data management framework that ensures seamless data accessibility and facilitates informed, timely decision-making across the organization.

Through these targeted objectives, Acme Inc. endeavors to construct a cohesive, data-centric infrastructure that not only enhances operational efficiency but also amplifies its market presence and growth trajectory.

Potential pitfalls in integrating a warehouse system with existing IT systems.

Integrating a warehouse system with existing IT systems at Acme Inc. presents several potential pitfalls that must be carefully considered. Firstly, **data inconsistency and poor quality** could arise due to the diverse data formats and standards used across the company's various systems (Andiyappillai 2020). This is evident in their challenges with inaccurate forecast estimates, delayed shipments, and damaged items, which may stem from discrepancies in how data is stored and managed.

- **Compatibility issues** are another concern, given Acme's heterogeneous IT landscape with different relational database products and versions. Communication problems and data-sharing issues between these disparate systems can arise due to migration towards a new warehouse management system (Andiyappillai 2020).
- **Complexity** is significant, especially considering Acme's plan to implement a data warehouse to meet analytic requirements. Without a clearly defined scope, they may encounter challenges with scope creep (Andiyappillai 2020), where additional requirements are added during the project, leading to delays and increased costs.
- **Data security risks** are heightened during integration, especially considering the sensitive customer and inventory data involved (Andiyappillai 2020). Acme must

ensure robust security measures are in place to protect against unauthorized access or breaches.

- **Performance issues** could arise if the new warehouse system is not optimized for Acme's existing IT infrastructure. This could result in slower response times and system crashes, impacting warehouse operations and data analysis (Andiyappillai 2020).

Lastly, **the lack of internal expertise** in integrating systems could also pose challenges for Acme. They may need to invest in training or seek external expertise to navigate the complexities of integration effectively.

Available Design Considerations and Which is the Best Option suits Acme Inc.

	Data Warehouse (Traditional Relational Database)	**Data Lake (Hadoop, AWS S3, Azure Data Lake Storage)**	**Graph Database (Neo4j)**	**Hybrid approach**	**Cloud-Based Solutions (AWS Redshift, Google BigQuery, Azure Synapse Analytics, MongoDB Atlas)**
Description	Centralized for structured data.	Centralized for diverse data types, supports real-time analytics.	Optimized for highly connected data.	Combines different approaches.	Scalable managed services on clouds.
Suitability	Structured data, specific analytics.	Diverse data, various analytics.	Relationship analysis, diverse data needs.	Organizations with varied requirements.	Scalable, cost-effective solutions.
Consideration	Scalability, upfront investment.	Data governance, complexity.	Relationship-based analysis, integration.	Integration, interoperability, security.	Scalability, cost management.
CAPEX	Higher due to infrastructure.	Lower with pay-as-you-go cloud services.	Moderate depending on deployment.	Varies based on chosen technologies.	Lower with usage-based pricing.
OPEX	Lower ongoing maintenance costs.	Higher due to usage-based pricing.	Moderate, includes maintenance.	Varies depending on technologies.	Based on usage, resource costs.
Scalability	Limited by hardware.	Unparalleled scalability.	Optimized for scaling horizontally.	Combines elements for scalability.	On-demand with elastic scaling.
Performance	Optimized for structured data.	Depends on processing, data volume.	High performance for graph-based queries.	Performance varies with integration.	High performance with distributed processing.

(Source: Herden 2020; Vukotic 2015; Swoyer 2021)

Few design considerations were found from different studies. Among them all, it is suggested that Acme Inc goes with the **Hybrid Approach**. Firstly, the company has diverse type of data, consisting of structured, semi-structured, and unstructured data. Utilizing the combination of traditional relational database and data lake can address this diversity. Secondly, using Graph database, i.e. Neo4j, benefits Acme Inc in a way to provide more advanced data analytics for analyzing relationships among entities, where it will be profitable in terms of forecasting product demands, solving customer complaints and item damage prevention. Thirdly, in terms of scalability, by embracing the hybrid approach, Acme can scale up the data management level to personalized used of cases, which leads to better targeted solution (Kondylakis et al. 2015). Fourthly, in terms of budgeting, the hybrid approach will have moderate amount of Capital Expenditure (CAPEX) and Operational Expenditure (OPEX). This cost management will be more efficient compared to the traditional and cloud-based solutions. Lastly, in terms of operational compliance, this hybrid approach allows better data integration and offers more flexibility to manage huge, diverse data types available within the existing IT system and even after new system integration.

Sizing/Estimation and Hardware Infrastructure

Infrastructure planning, particularly sizing and estimating hardware requirements, is crucial for the successful implementation of a data warehousing solution at Acme Inc. (Nagabhushana, 2006). Given the heterogeneous nature of the company's IT systems, this process becomes even more challenging. Here's how Acme Inc can approach sizing and estimating their hardware infrastructure:

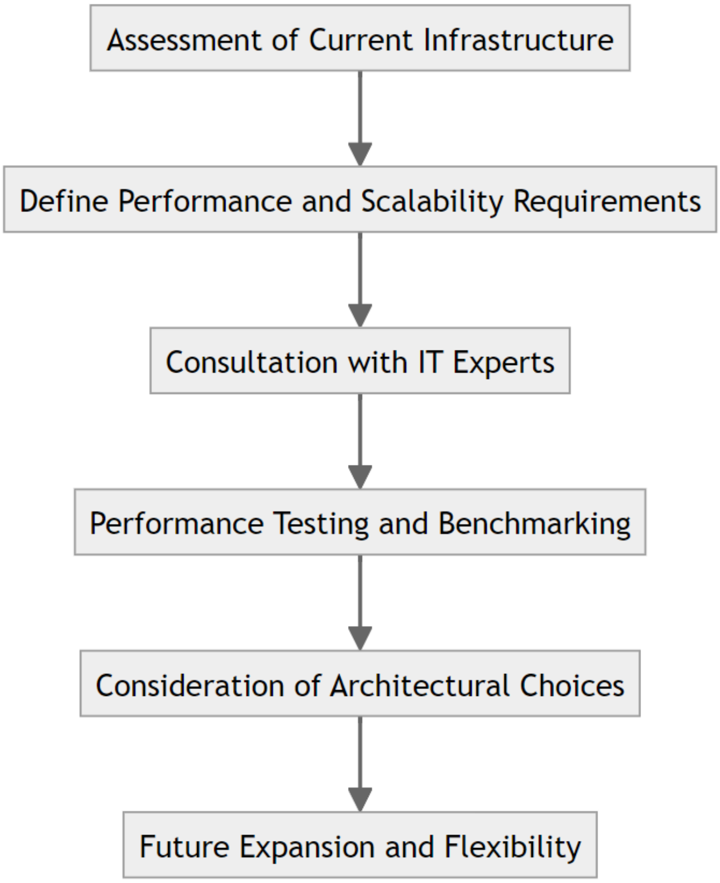


Figure 1:Acme sizing and estimating hardware approach.

1. Assessment of Current Infrastructure: Before determining hardware requirements, Acme needs to assess its current infrastructure thoroughly. This includes evaluating existing servers, storage systems, network bandwidth, and database servers used across various units. Understanding the capacity, performance, and limitations of the current infrastructure is essential for planning upgrades or additions.

2. Define Performance and Scalability Requirements: Acme should identify the performance metrics and scalability requirements for the proposed data warehousing solution (Krishnamoorthy, 2015). This includes factors such as the volume of data to be stored and processed, the number of concurrent users accessing the system, and expected growth in data over time. By defining these requirements, Acme can ensure that the hardware infrastructure can support current and future needs.

3. Consultation with IT Experts: Given the complexity of the infrastructure and the importance of accurate sizing, Acme should consult with IT experts or engage with external consultants specialising in data warehousing and infrastructure planning. These experts can provide valuable insights into best practices, industry standards, and emerging technologies relevant to hardware infrastructure for data warehousing.

4. Performance Testing and Benchmarking: Acme should conduct performance testing and benchmarking exercises to validate hardware requirements. This involves simulating realistic workloads and analysing system performance under different scenarios. By testing various configurations and hardware components, Acme can identify optimal setups that meet performance targets and scalability needs (Campbell, 2024).

5. Consideration of Architectural Choices: Acme must align hardware sizing with chosen architectural choices for the data warehousing solution. For example, if they opt for a distributed architecture with multiple data marts or a centralised architecture with a single, large data warehouse, the hardware requirements will vary significantly. Each architectural choice has implications for hardware configuration, such as storage capacity, processing power, and network bandwidth (Armstrong, 1977).

6. Future Expansion and Flexibility: Acme should factor in future expansion and flexibility while sizing hardware infrastructure. This includes considering potential increases in data volume, changes in usage patterns, and evolving business requirements. Building scalability into the infrastructure design ensures that Acme can adapt to growth and changes in the future without significant disruptions or overhauls (Fienberg, 2006).

By following these steps and considering the unique challenges posed by their heterogeneous IT environment, Acme can effectively size and estimate hardware requirements for their data warehousing solution, laying a solid foundation for decision support infrastructure.

Change Management and Communication Plan:

Changing management plays a crucial role in ensuring the successful implementation of the data warehousing solution at Acme Inc. Key strategies include:

- **Stakeholder Engagement:** Engage stakeholders at all levels of the organisation to gain their support and involvement in the change process. Communicate the benefits of the data warehousing initiative and address any concerns or resistance proactively (Hassin, 2018).
- **Training and Skill Development:** Provide comprehensive training programs to equip employees with the necessary skills and knowledge to use the new data warehousing system effectively. Offer ongoing support and resources to facilitate learning and adoption.
- **Change Champions:** Identify and empower change champions within the organisation to advocate for the data warehousing initiative and drive adoption among their peers. These individuals can serve as role models and mentors to inspire others to embrace the change.
- **Communication Plan:** Develop a communication strategy to keep all stakeholders informed and engaged throughout the implementation process. Utilise various communication channels such as emails, newsletters, intranet portals, and town hall meetings to provide regular updates, address concerns, and celebrate milestones (Hassin, 2018).

By implementing these change management strategies and effective communication plans, Acme Inc can minimise resistance to change, foster a culture of collaboration and innovation, and ensure a smooth transition to the new data warehousing solution.

Warehouse Management

Improving Forecasting Accuracy

The challenge of forecast inaccuracy in Acme Inc., especially with deviations as significant as 30-40% for some SKUs, necessitates a multifaceted approach to enhance the precision of inventory forecasts. To tackle this issue, Acme can integrate the following data sources:

1. Transaction Data:
    a. Point-of-Sale (POS) data with SKU-level sales quantities and revenue.
    b. Product returns/refunds data

c. Inventory data from stores and warehouses.

2. Product Data:
    a. Product master with details like category, brand, specifications
    b. Product hierarchies and associations
3. Customer Data:
    a. Customer profiles (demographics, psychographics)
    b. Customer loyalty program data
    c. Customer segmentation data
4. Promotions Data:
    a. Promotion details like type, discount %, start/end dates.
    b. Promotion performance data
5. Store Data:
    a. Store locations, characteristics (size, format, etc.)
    b. Store clusters/regions.
    c. Store planograms and merchandizing data.
6. External Data:
    a. Economic indicators like GDP, consumer confidence
    b. Weather/climate data
    c. Events/holiday calendars
    d. Competition and market share data

Once this granular multi-dimensional data is integrated in the data warehouse, Acme can build sophisticated forecasting models leveraging techniques like a blend of historical sales data analysis, market trend evaluation, and advanced predictive analytics techniques (Edward & Sabharwal 2015). A more granular approach to forecasting that incorporates machine learning models can analyze patterns and trends from past sales data, seasonal fluctuations, and market dynamics to predict future demand with higher accuracy.

Advanced Forecasting Techniques:

- Time Series Models:
  - ARIMA, Exponential Smoothing with Box-Cox transformations
  - Seasonal ARIMA (SARIMA) to handle seasonality patterns.
  - Additive/Multiplicative decomposition for trend/cycle components
- Regression Models:
  - Multi-variable regression to quantify impact of promotions, economic factors.
  - LASSO/Ridge/Elastic Net regularization for variable selection.
- Machine Learning Models:
  - Decision Trees (Random Forests) to handle non-linear relationships.
  - Gradient Boosting Machines for accurate combined models
  - Neural Networks like LSTMs for sequence/temporal learning
  - Clustering for product/store/customer segments
- Bayesian Forecasting with prior belief updating as new evidence arrives.
- Ensemble/Combined forecasts from multiple models for best accuracy

Additionally, integrating external factors such as economic indicators, competitor activities, and social media sentiment analysis can provide a more comprehensive view of potential demand influences (Edward & Sabharwal 2015). This approach necessitates the consolidation of data from various sources into a centralized data warehouse, enabling more sophisticated data mining and analytics processes. By continuously refining these models with new data, Acme can enhance the adaptability and accuracy of its forecasts over time.

Addressing Delayed Shipments

To mitigate issues related to delayed shipments and enhance supply chain efficiency, the following integrated supply chain analytics solution is proposed:

1. **Data Extraction:**
   a. Consolidate data from warehouse management systems, transportation systems, supplier systems, and ERP/order data into the data warehouse.
2. **Real-time Visibility Dashboards:**
   a. Build dashboards to provide real-time visibility into order statuses, expected ship dates from warehouses, shipment tracking from transporters, and estimated delivery dates to retail stores.
3. **Historical Data Analysis:**
   a. Analyze historical shipment and order data to identify common bottlenecks in warehouse operations, frequent delays from specific transporters or routes, and issues with supplier lead times.
4. **Predictive Analytics:**
   a. Apply predictive analytics to forecast potential delays based on current orders, shipments, and transporter performance data.
   b. Use these forecasts to proactively address potential delays through route optimization and carrier selection (Arora & Gosain 2020)
5. **Process Optimization:**
   a. Optimize transportation routes, modes, and carrier selection based on analysis to minimize delays, prioritizing urgent deliveries.
   b. Implement supplier performance tracking against lead times and order fulfillment SLAs.
6. **Collaborative Supply Chain Visibility:**
   a. Provide retailers with collaborative supply chain visibility, including delay alerts and root cause analysis, to enhance transparency and trust.

Architecture for Warehouse Estimation

To increase warehouse operational efficiency and address the challenges of forecast inaccuracies and delayed shipments, Acme Inc. needs to adopt a comprehensive architectural framework that supports real-time data integration, advanced analytics, and seamless communication across the supply chain (Mourya 2012). The proposed architecture comprises three core components:

- **Data Integration Layer:** Centralizes data from various internal and external sources, including ERP systems, CRM, market data, and social media, into a unified data warehouse (Armstrong 1997). This layer uses ETL (Extract, Transform, Load) processes to ensure data quality and consistency.
- **Analytics and Forecasting Engine:** Utilizes advanced machine learning models and predictive analytics tools to process the integrated data for demand forecasting, inventory optimization, and route optimization (Campbell et al. 2017). This engine should be capable of handling complex algorithms and delivering actionable insights in real-time.
- **Supply Chain Visibility Platform:** A dashboard that provides end-to-end visibility of the supply chain, offering real-time tracking of inventory levels, shipments, and delivery status (Bhatia 2019). This platform should facilitate communication between Acme, its logistics partners, and retailers, enabling more efficient coordination and rapid response to potential issues.

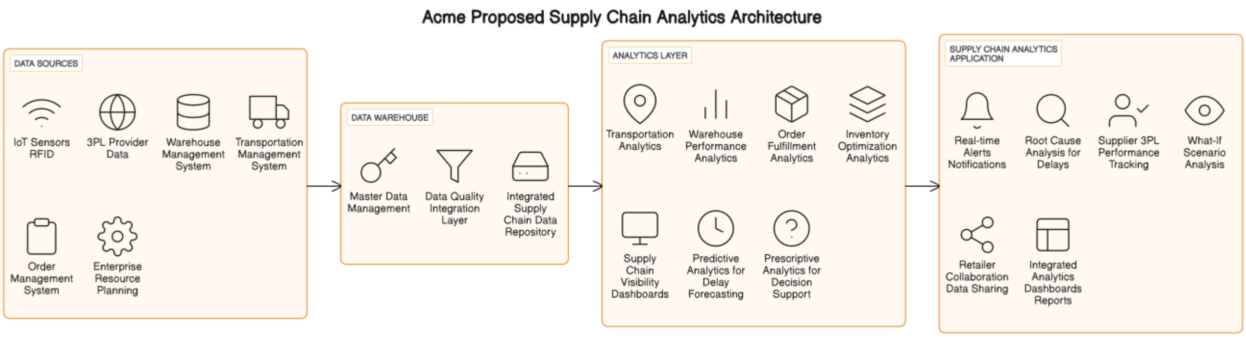


Figure 2:Acme Proposed supply Chain Analytics Architecture

The diagram shows Acme Inc.'s proposed supply chain analytics architecture, which consists of four key components:

- Data Sources: Collects data from IoT sensors, 3PL providers, and various management systems.

- Data Warehouse: Centralizes and refines the data through master data management and integration layers.
- Analytics Layer: Processes the data to provide insights on transportation, warehouse performance, order fulfillment, and inventory.
- Supply Chain Analytics Application: Utilizes the insights to deliver real-time alerts, perform root cause analysis, track supplier performance, and facilitate retailer collaboration.

This architecture is designed to streamline Acme's supply chain, improve efficiency, and enhance decision-making capabilities.

Product promotion Solution

Acme needs to first create a centralized data warehouse to consolidate sales data from all stores, then use ETL (Extract, Transform, Load) processes to extract data from each store's POS and ERP systems to convert it into a standardized format, and load it into the data warehouse from these disparate sources. A data warehouse platform solution that is scalable, flexible and cost-effective is proposed (A.Gupta et al., 2020).

Secondly, design Multi-dimensional data models that accommodate slicing and dicing of sales data by various dimensions such as product category, store location, time period, customer segment (based on purchasing behaviour, demographics and preferences), and promotional activity. Subsequently, use a star schema because of its simplicity, performance for slicing and dicing sales data, denormalization and ease with navigating and analyzing data from the dimension tables to the fact table (Bhatia 2019).

Thirdly, for data analysis, implement OLAP (Online Analytical Processing) tools to facilitate interactive data analysis allowing users to slice, dice, drill-down and pivot data across different dimensions and exploration of sales data from different perspectives. Develop dashboard, reports, and ad-hoc querying capabilities that provide insights into sales performance, promotional effectiveness, and customer behaviour (Prabhu 2007). This will also enable business users to perform self-service analytics. Incorporate data visualization techniques to present insights in an easy-to-understand format. Use advanced analytics techniques such as predictive modeling and segmentation analysis to identify trends, patterns, and correlations in the data.

Moving on**,** apply advanced analytics techniques such as predictive modeling, clustering, and association analysis to uncover hidden patterns and trends in the data (Mourya 2012). Build predictive models to identify customer segments with the highest propensity to respond to promotions and loyalty programs including the right customer targets for promotional campaigns and customer churn. Use data mining algorithms to analyze historical sales data and identify patterns of customer behaviour, preferences, and purchase patterns.

- Train the models using historical sales data, customer demographics, and transactional behaviour.
- Deploy the models to generate targeted marketing recommendations and personalized offers for individual customers.

Subsequently, define key performance indicators (KPIs) to measure the effectiveness of promotional campaigns, including Return on Investment (ROI), customer acquisition cost, customer lifetime value, and promotional lift. Develop tracking mechanisms for real-time monitoring and evaluate the impact of promotions on sales revenue, customer retention, and overall profitability. Use A/B testing and experimental design methodologies to compare different promotional strategies and optimize marketing spend (Bhatia 2019).

Also, implement a customer relationship management (CRM) system to manage the loyalty card program and track customer interactions. Integrate the CRM system with the data warehouse to leverage customer data for personalized marketing communication and loyalty initiatives. Use analytics to measure the impact of the loyalty card program on customer retention and overall profitability (Mourya 2012).

Lastly, initiate a pilot phase to create awareness, provide training and support to ensure that users across the organization understand how to leverage the data and analytics tools effectively. Encourage a data-driven culture by promoting the value of analytics. Establish governance processes and data stewardship responsibilities to maintain data quality, integrity, and security. Continuously improve on the system based on feedback, changing business requirements, and technological advancement (Prabhu 2007).

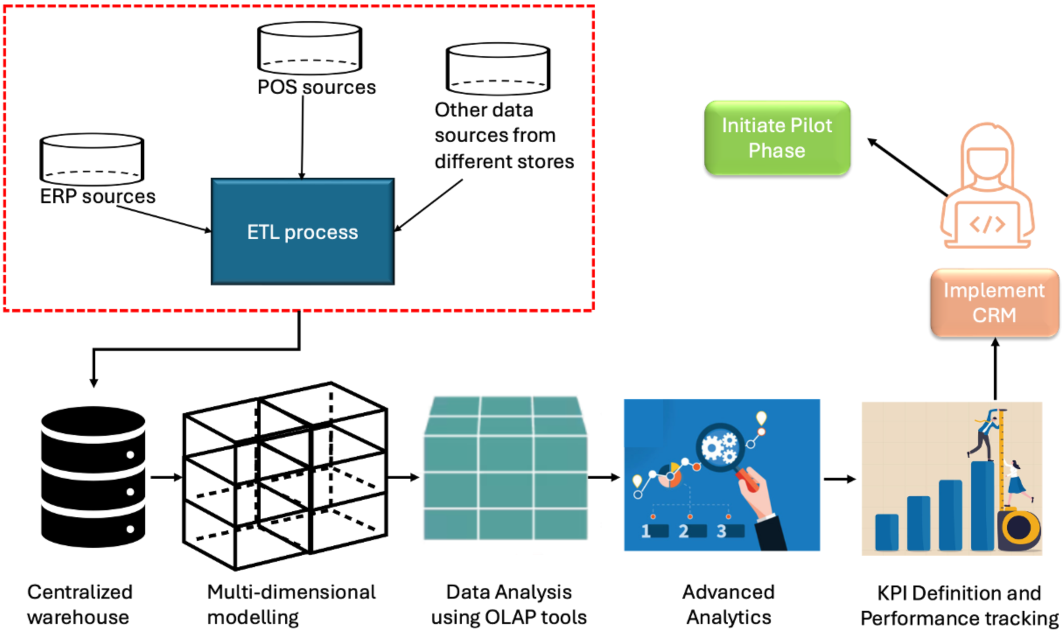


Figure 3:Acme Production promotion solution.

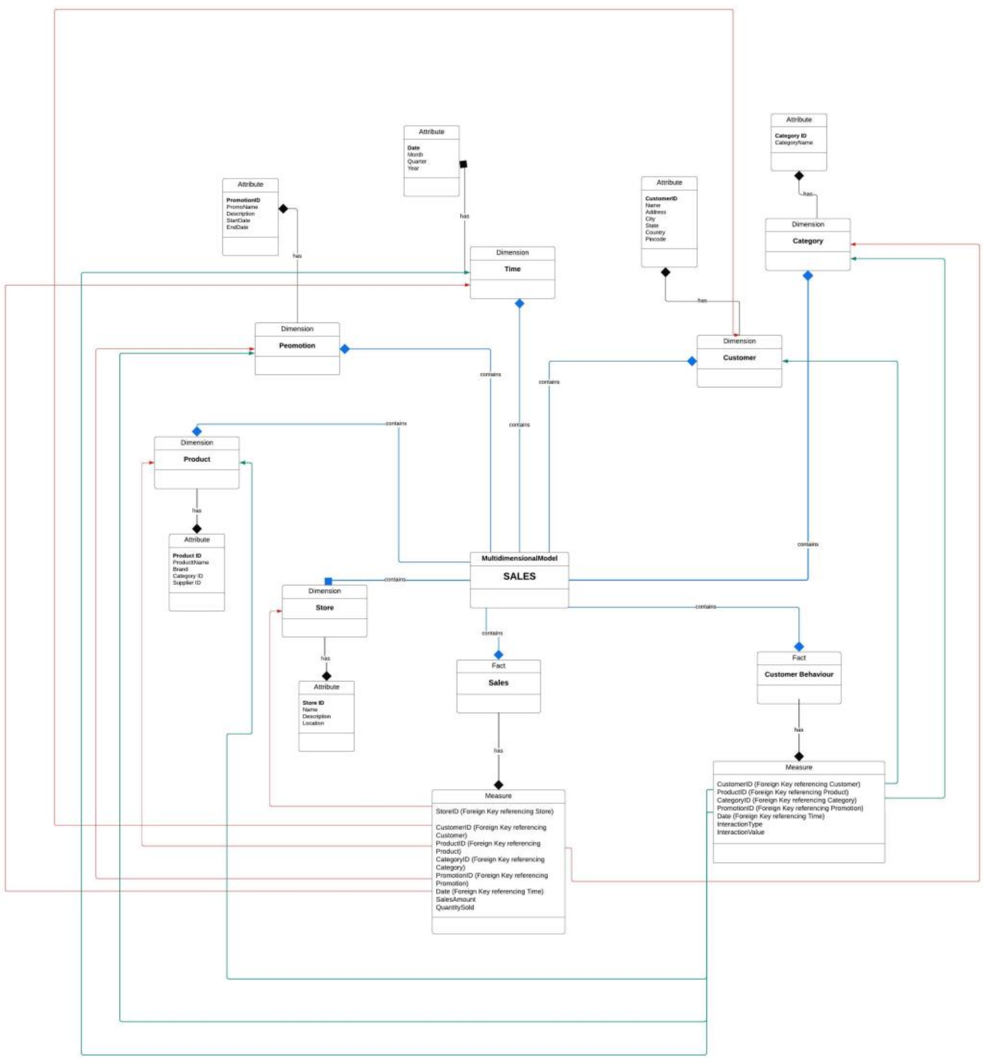


Figure 4:Acme Multi-dimensional model.

Data Warehouse Architecture Evaluation

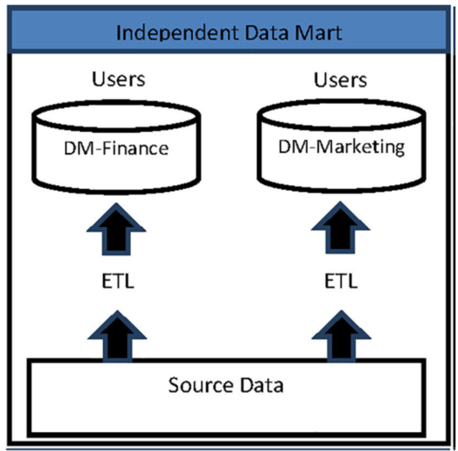
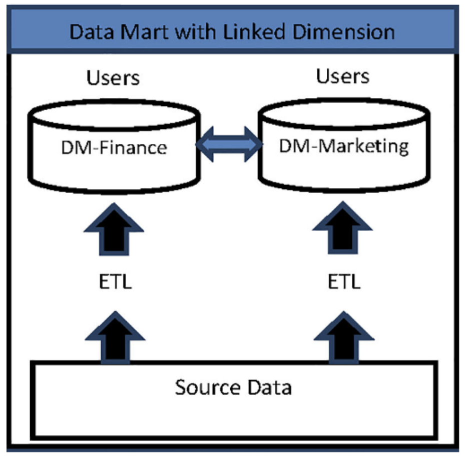
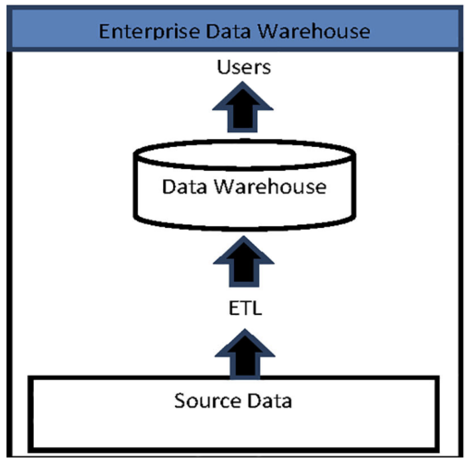
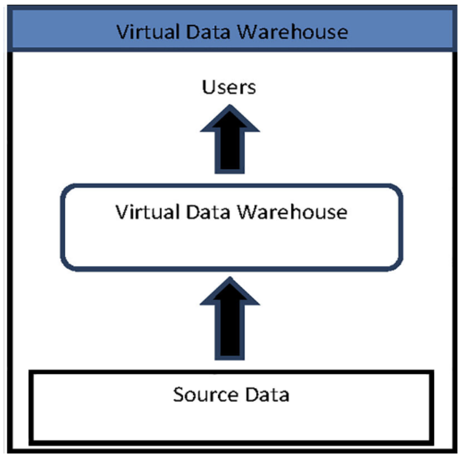
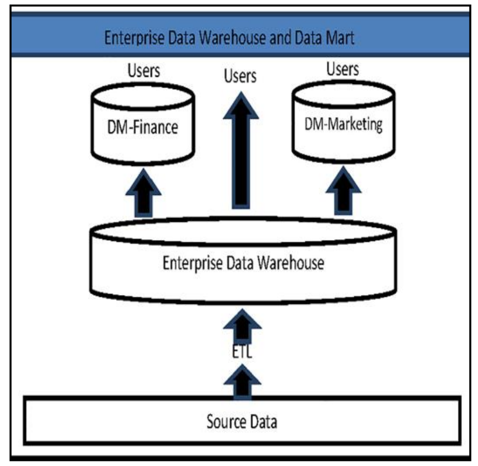
There are various data warehouse architectures as shown below:

Architecture Type	Description
Single-Layer	Virtual data warehouse implemented as a multidimensional view of operational data; lacks separation between analytical and transactional data processing.
Two-Layer	Separates data sources and data warehouse layers, facilitating distinction between analytical and transactional data processing; consists of source layer, data staging, data warehouse layer, and analysis.
Three-Layer	Physically implements source layer, reconciled layer, and data warehouse layer; ensures integrated, consistent, accurate, and detailed data after integrating and cleansing source data.
Independent Data Marts	Consists of separate, inconsistent data marts, making data analysis difficult; often replaced for better data integration.
Bus	Logically integrates data marts with an enterprise-wide view of information; similar to independent data marts architecture.
Hub-and-Spoke	Central enterprise data warehouse (hub) feeding data marts (spokes); emphasizes scalability, extensibility, and retrieving large amounts of information.
Centralized	Single centralized data warehouse containing integrated data and no dependent data marts; recommended by Bill Inmon as a specific implementation of hub-and-spoke architecture.
Distributed	Integrates existing data warehouses/data marts to provide a single solution; logically or physically integrates each data warehouse/data mart using joint keys, global metadata, distributed queries, and other methods.

(Source: Blažić et al., 2017)

Based on Exhibit 6 from the case study, the differences among the data warehouse alternatives are itemized below:

Differences among these alternatives

	Independent Data Mart	**Data Mart with Linked Dimension**	**Enterprise Data Warehouse (EDW)**	**Virtual Data Warehouse**	**Enterprise Data Warehouse and Data Mart**
					
Pros	**Tailored to specific business unit needs:** Each Data Mart can be customized to meet the unique analytical requirements of different departments. **Autonomy:** Business units have control over their own data and analysis, allowing for flexibility and agility in decision-making. **Simplified management:** Each Data Mart operates independently, making it easier to manage and maintain.	**Data Consistency:** Shared dimensions or reference data facilitates consistency across Data Marts, enabling more accurate and integrated analysis. **Cross functional analysis:** The linkage between Data Marts allows for analysis that spans multiple business units to provide deeper insights. **Departmental autonomy:** Despite the linkage, each Data Mart retains autonomy, allowing for customization to specific departmental needs.	**Centralized data repository:** The EDW provides a single source of truth for all integrated data, ensuring consistency and reliability. **Unified view:** Users across the organization have access to a consistent and comprehensive view of data, enabling informed decision making. **Scalability:** Ability to scale to accommodate growing data volumes, analytical needs which supports the Acme Inc's long-term growth.	**Agility and Flexibility:** The virtual layer allows for on-demand access to data from different sources enabling agile and flexible analysis. **Cost-effective:** Avoids the need for upfront investment and storages, potentially reducing costs associated with data warehousing. **Integration Simplicity:** It provides a logical layer that extracts underlying data sources, reducing complexity.	**Combination of centralized and decentralized:** It provides the benefits of a centralized EDW while allowing for customized analytics withing individual business units. **Data Consistency and autonomy:** It ensures consistence across the organization while allowing for departmental autonomy and customization. **Scalability and flexibility:** It can scale to accommodate growing data volumes and analytical needs of acme inc.

Cons	**Data Silos:** This limits cross-functional analysis and can lead to inconsistencies. **Redundancy:** There could be duplicated data across multiple Data Marts, leading to inefficiencies in storage and maintenance. **Lack of enterprise-wide view:** Acme inc will struggle to achieve a unified view of data across departments therefore hindering holistic informed decision-making.	**Complexity:** Managing and maintaining linked dimensions requires careful coordination and governance to ensure consistence, integrity and accuracy. **Redundancy:** There could still be redundancy in data storage and maintenance. **Integration challenges:** Linking Data Marts may pose challenges in terms of data integration and synchronization as Acme inc expands.	**Upfront investment:** The cost of building and implementing an EDW in terms of infrastructure, technology, and resources is significant. **Complexity:** Designing and managing a centralized EDW is complex and requires careful planning, governance, and coordination across departments. **Performance bottlenecks:** With large volumes of concurrent queries, EDW may face performance issues as a central repository for all data.	**Performance Limitations:** Query performance may be impacted by the need to access data from different sources in real-time. **Data governance challenges:** Maintaining data quality, consistency, and security can be challenging in a virtual environment with multiple data sources. **Dependency on source systems:** It relies on the availability and reliability of underlying data sources which may introduce risks and dependencies.	**Complexity:** The architecture becomes more complex whilst combining both centralized and decentralized approaches. Therefore, it requires careful design and management. **Integration Challenges:** Coordinating data flow between the EDW and Data Marts may pose integration challenges, especially as Acme inc grows. **Maintenance overhead:** It requires ongoing maintenance and governance to ensure consistency, integrity, accuracy, and efficiency across the EDW and Data Marts.

(Source: Blažić et al., 2017; Sun et al., 2013)

Best benchmarks/criteria to select the appropriate and the best Data Mart/ Warehousing Architecture

1. **Business Requirements**: Analyze current and future business needs including reporting, analysis, and decision-making requirements to ensure alignment with strategic objectives and desired insights (Inmon et al., 2014).
2. **Scalability**: Assess the architecture's ability to grow with data volumes and user demands and analytical complexity over time while maintaining performance (Qian et al., 2009)
3. **Flexibility and Agility**: Evaluate the architecture's ability to adapt to changing business and technological landscapes to support agile decision-making (Qian et al., 2009).
4. **Data Integration and Quality**: Ensure seamless integration of disparate data sources while maintaining data quality and governance (Inmon et al., 2014).
5. **Performance and Accessibility**: Evaluate query response times and accessibility across the organization, prioritizing user experience (Qian et al., 2009).
6. **Security and Compliance**: Consider robust security measures and compliance with regulations to protect sensitive data. Consider the architecture's security features and capabilities, including data encryption, access controls, and compliance with regulatory requirements (e.g., GDPR, HIPAA) (Dias et al.,2008).
7. **Cost and ROI**: Assess total ownership costs and potential returns on investment, considering both initial and ongoing expenses (Inmon et al., 2014).
8. **Ease of Management and Maintenance**: Evaluate manageability and maintenance requirements, seeking tools to streamline processes (Dias et al.,2008).
9. **Organizational Readiness**: Assess readiness for implementation and support, including technical skills and alignment with organizational goals (Inmon et al., 2014).
10. **Industry Best Practices and Standards**: Benchmark against established methodologies and seek expert guidance for validation and optimization (Dias et al.,2008).

Best Option that suits Acme Inc

Given Acme Inc's heterogeneous IT landscape and diverse business requirements, the most suitable architecture would be a combination of **Enterprise Data Warehouse (EDW) and Data Mart**, for the below reasons:

1. **Integration and Centralization:** Acme Inc requires integration of data from multiple sources, including ERP systems, CRM, and operational databases. An Enterprise Data Warehouse (EDW) serves as a centralized repository for all integrated data, ensuring consistency and reliability.
2. **Scalability:** As Acme Inc continues to expand its operations and data volume, an EDW can scale to accommodate growing data needs. Additionally, the use of Data Marts allows for the creation of specialized subsets of data tailored to specific business units or departments (e.g., Finance, Marketing), providing scalability and flexibility.
3. **Analytical Capabilities:** This architecture supports the diverse analytical requirements of Acme's business units, such as analyzing sales performance, inventory trends, and customer behaviour.
4. **Data Consistency and Accuracy:** By centralizing data in an EDW and distributing relevant subsets to Data Marts, Acme can ensure data consistency and accuracy across the organization. This is essential for making informed business decisions based on reliable data.
5. **Flexibility:** Each business unit can access the relevant subset of data stored in their respective Data Mart, while still benefiting from the centralized data management provided by the EDW.
6. **Cost Efficiency:** While implementing and maintaining an EDW and Data Marts may involve initial investment, the long-term cost efficiency can be significant. By consolidating data management and analytics infrastructure, Acme can reduce redundancy and streamline operations.

Cloud-based Architecture pros and Cons

Pros:

1. **Scalability:** Cloud-based data warehouses can scale up or down dynamically based on demand, allowing organizations to easily accommodate fluctuations in data volumes and user loads without the need for extensive hardware investments (Arora 2021; (Swoyer 2021).
2. **Flexibility:** Cloud platforms offer a wide range of data storage and processing services, allowing organizations to choose the tools and technologies that best fit their needs. This flexibility enables rapid innovation and experimentation with new data sources, analytics tools, and techniques (Swoyer 2021).
3. **Cost Efficiency:** Cloud-based data warehouses eliminate the need for upfront hardware procurement and infrastructure investments, reducing capital expenditures (CapEx). Instead, organizations pay for cloud resources on a pay-as-you-go or subscription basis, leading to more predictable operational expenses (OpEx) (Swoyer 2021).
4. **Global Accessibility:** Data can be accessed from anywhere with an internet connection, promoting collaboration and remote work including simplifying disaster recovery/business continuity (Swoyer 2021).
5. **Managed Services:** Cloud providers offer fully managed data warehouse services that handle routine tasks such as hardware provisioning, software updates, security patching, and backup and recovery. This offloads administrative burden from internal IT teams and allows them to focus on higher-value activities (Swoyer 2021).
6. **Integration Capabilities:** Cloud platforms provide robust integration capabilities, allowing organizations to easily connect their data warehouse with other cloud services, data sources, and analytics tools. This facilitates data sharing, interoperability, and cross-functional collaboration (Swoyer 2021).

Cons:

1. **Data Security and Privacy:** Storing sensitive data in the cloud raises concerns about data security, privacy, and compliance with regulatory requirements. Organizations must carefully evaluate the security features and certifications offered by cloud providers and implement appropriate encryption, access controls, and compliance measures (Arora 2021).
2. **Performance and Latency:** Cloud-based data warehouses may experience performance issues and latency, especially when dealing with large volumes of data or

complex analytical queries. Organizations need to optimize data transfer, processing, and storage to minimize latency and ensure acceptable performance levels (Arora 2021).

3. **Vendor Lock-In:** Adopting a cloud-based data warehouse may lead to vendor lock-in, where organizations become dependent on a specific cloud provider's ecosystem, tools, and pricing models. Migrating data and workloads between cloud platforms can be complex and costly, limiting flexibility and agility (Arora 2021).
4. **Data Transfer Costs:** Transferring large volumes of data between on-premises systems and the cloud, or between different cloud regions or providers, can incur significant data transfer costs. Organizations need to carefully manage data movement and optimize network bandwidth usage to control expenses (Swoyer 2021).
5. **Compliance and Legal Issues:** Cloud-based data warehousing raises compliance and legal concerns related to data sovereignty, jurisdictional regulations, and cross-border data transfers. Organizations must ensure compliance with industry-specific regulations (e.g., GDPR, HIPAA) and contractual obligations when storing and processing data in the cloud (Swoyer 2021).

Neo4J based solution

Based on the provided information and the requirements outlined, an architecture for a Neo4j-based solution for Acme Inc was generated:

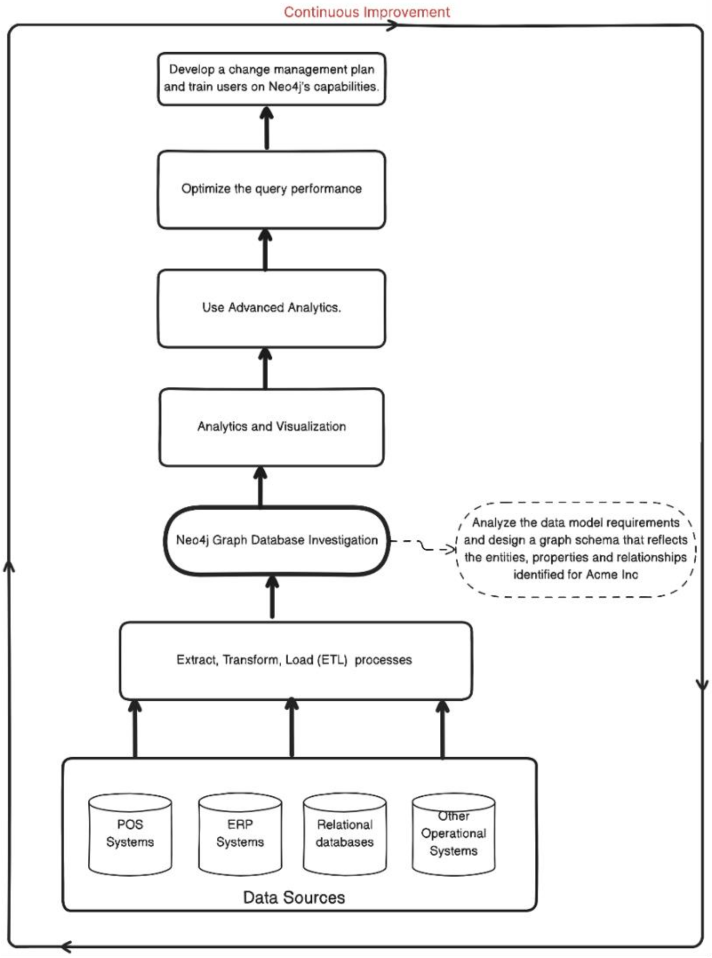


Figure 5: Neo4j solution

The below steps are to be followed:

1. Setup the Neo4j database environment either locally or on a cloud platform like Neo4j Aura. Connect to the Neo4j database using tools like Neo4j Browser or Neo4j Desktop.

2. For the Architecture, proceed with the ETL (Extract, Transform, Load) Process. Extract data from various sources such as POS systems, ERP systems, relational databases, and other operational systems. Transform and clean the extracted data to ensure consistency and quality (Vukotic 2015).

3. Following that, load the transformed data into the Neo4j graph database using Neo4j's import tools or custom script. Design a graph data model representing entities such as

products, stores, customers, promotions, orders, shipments, and activities in the order fulfillment pipeline. Define nodes and relationships to represent these entities and their attributes.

4. Subsequently, Utilize Cypher queries to perform multi-dimensional data analysis, slicing, dicing, and drill-down operations. Develop custom queries to calculate key metrics such as sales revenue, inventory levels, and promotional effectiveness (Kemper 2015).
5. Proceed to use Neo4j's graph visualization capabilities to create interactive dashboards and visualizations in a clear and understandable for stakeholders:
6. Warehouse Management interface for visualizing and managing inventory data in a graph format which can track inventory levels, monitor shipments, and optimize warehouse operations.
7. Sales analysis interphase to generate reports, track sales performance, identify insights and trends, and promotional effectiveness.
8. Customer targeting interface to equip the relevant teams with tools for customer segmenting, targeted promotions, marketing campaigns or loyalty programs.
9. Use Neo4j's Graph Data Science library to perform advanced analytics tasks such as graph algorithms (e.g., community detection, centrality analysis, and pathfinding) (Vukotic 2015).
10. Lastly, continuously iterate on the solution based on feedback from stakeholders and evolving business requirements. Fine-tune the data model, analytical queries, and visualization techniques to improve the effectiveness and usability of the solution over time (Kemper 2015).

Based on the above, the below was implemented on Neo4J browser using Arbitrary data based on Exhibit 2 of the Acme Case study:

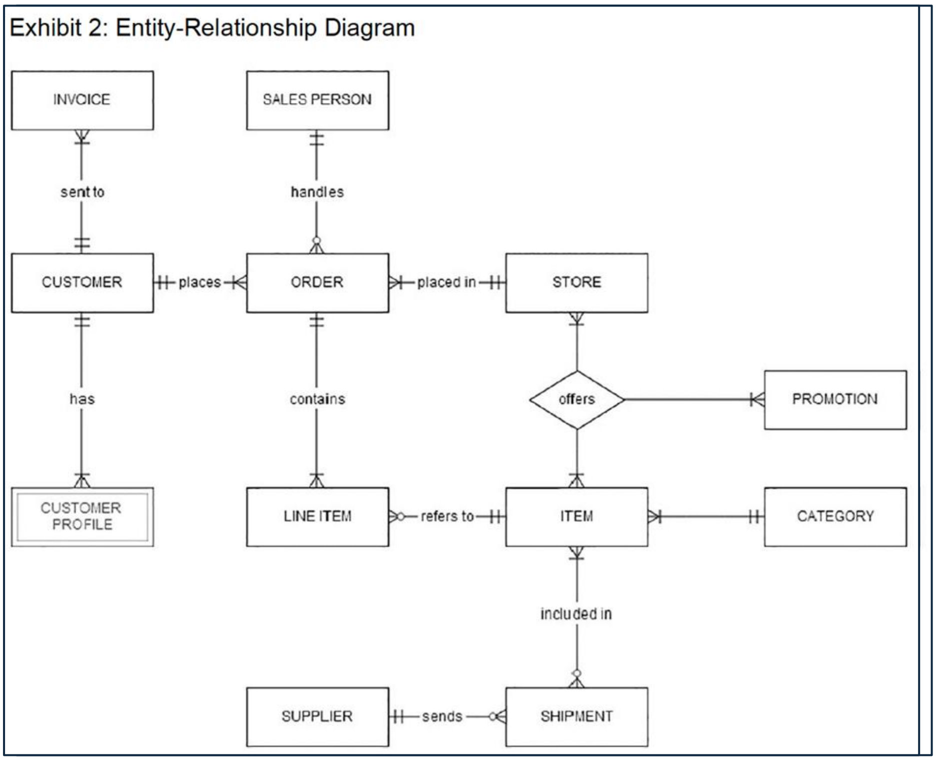


Figure 6:ERD Diagram for Related Entities for Acme Inc. Case Study

Server version	Neo4j/5.17.0
Server address	localhost:7687
Query	CREATE (customer:Customer { customerId: '550', name: 'Rita Edemu', address: '4 Gordon Groove', city: 'South Yarra', state: 'Victoria', country: 'Australia', pincode: '314156' }) CREATE (customer_profile:CustomerProfile { customerId: '550', demographicProfile: 'Demographic Profile', psychographicProfile: 'Psychographic Profile' }) CREATE (customer)-[:HAS]->(customer_profile) CREATE (customer_order:Order { orderId: '100', orderDate: date('2024-04-01'), salesTax: 10.00, discount: 5.00, freight: 3.00, amount: 100.00 }) CREATE (customer)-[:PLACES]->(customer_order) CREATE (line_item:LineItem { lineItemId: '202', name: 'Product Name', description: 'Product Description', price: 50.00 }) CREATE (customer_order)-[:CONTAINS]->(line_item) CREATE (invoice:Invoice { invoiceId: '789', customerName: 'Ravidu Abhinav', address: 'Address', amount: 500.00 }) CREATE (invoice)-[:SENT_TO]->(customer) CREATE (sales_person:SalesPerson { salesPersonId: '105', name: 'Kenny Ammie', designation: 'TeamLead', age: 30, dob: date('1994-01-01'), address: '370 Little Lonsdale' }) CREATE (sales_person)-[:HANDLES]->(customer_order) CREATE (store:Store { storeId: '202', name: 'Acme Inc', description: 'The Good guys', location: 'GroundFloor' }) CREATE (store)-[:PLACED_IN]->(customer_order) CREATE (store)-[:OFFERS]->(line_item) CREATE (item:Item { itemId: '101', productName: 'PlayStation5', brand: 'Sony', category: 'Games', department: 'Electronics' }) CREATE (store)-[:OFFERS]->(item) CREATE (promotion:Promotion { promotionId: '201', promoName: 'Everyday Rewards', description: 'Jumbo Pack', startDate: date('2024-04-01'), endDate: date('2024-04-30') }) CREATE (store)-[:OFFERS]->(promotion) CREATE (shipment:Shipment { shipmentId: '301', shipDate: date('2024-04-01'), deliveryDate: date('2024-04-02') }) CREATE (line_item)-[:REFERS_TO]->(item) CREATE (item)-[:INCLUDED_IN]->(shipment) CREATE (supplier:Supplier { supplierId: '401', supplierName: 'Harvey Norman', description: 'Electronics Section', location: 'Queensland' }) CREATE (supplier)-[:SENDS]->(shipment) CREATE (category:Category { categoryId: '501', categoryName: 'All Category', description: 'All Description' }) CREATE (category)-[:ITEM]->(item)

Figure 7:Arbitrary(Dummy) Data on Neo4j

Figure 8:Acme Inc. Neo4j with Dummy Data

Mongo-DB Based Solution

A framework was generated to address the issues encountered by Acme Inc., shown in the diagram below.

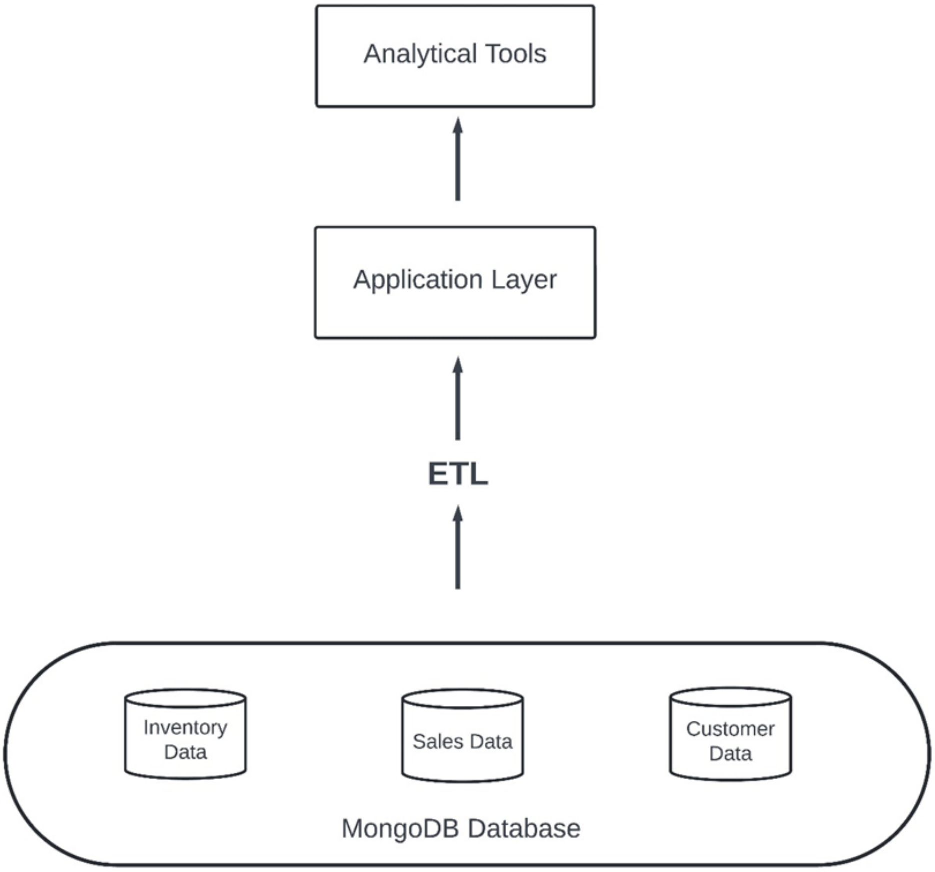


Figure 9:MongoDB-Based Framework

At the initial stage, the MongoDB database should be built in the MongoDB Atlas and connected with the MongoDB Compass. Not to mention, the inventory, sales, and customer data should be available inside the MongoDB environment to address each issue (warehouse management, product promotions, and customer targeting), respectively.

After that, the ETL Process (Extract, Transform, Load) is proceeded. Data is **extracted** from various sources such as ERPs (Enterprise Resource Planning systems) and relational databases. Extracted data is then **transformed** to clean and prepare it for analytics. This involves tasks such as data cleaning (removing duplicates, handling missing values), data normalization, and data enrichment (adding calculated fields and joining datasets) (Erraji, Maizate, Ouzzif 2022). Transformed data is then **loaded** into the respective MongoDB collections (Inventory Data, Sales Data, Customer Data) using appropriate scripts or tools. This ensures the data is structured and ready for analysis.

Once the ETL process is completed, the Application layer needs to be managed. Three interfaces should be created, detailed below:

- **Warehouse Management Interface:** this allows warehouse managers to view and manage inventory data. They can track stock levels, monitor incoming and outgoing shipments, and optimize warehouse operations (Edward & Sabharwal 2015).
- **Sales Analysis Interface:** this interface is for sales analysts to perform in-depth analysis of sales data. They can generate reports, track sales performance, identify trends, and make data-driven decisions for promotions and marketing strategies (Edward & Sabharwal 2015).
- **Customer Targeting Interface:** this interface provides tools for marketing and customer relationship teams. They can segment customers based on their profiles and purchase history, create targeted promotions or loyalty programs, and improve customer engagement (Edward & Sabharwal 2015).

In addition, APIs for Data Access are also provided for facilitating external system integration with MongoDB systems, allowing for seamless data sharing and real-time updates.

Finally, the analytical tools should be decided to analyze the data. The company can utilize the MongoDB aggregation technique for performing complex aggregations and analytics operations on the data stored in MongoDB collections. It allows for tasks such as grouping, filtering, and calculating aggregate values. In addition, Data mining Models can be generated via MongoDB's built-in machine learning model or integrated with external machine learning frameworks for advanced analytics (Botoeva et al. 2018). This step shall be highlighted to address the issues, such as predictive modeling for inaccurate stock calculation, clustering customers for loyalty card programs, and measuring promotional campaign effectiveness.

The ER Diagram presented in the Exhibit 2 has been successfully converted into the MongoDB database, as shown in ***Figure 10***. Furthermore, MongoDB Chart feature inside the MongoDB Atlas can also be utilized to create insightful visualization from the provided dataset. Two samples for data visualization using MongoDB Chart was obtained from the internet, as shown in ***Figure 11*** and ***Figure 12***, respectively.

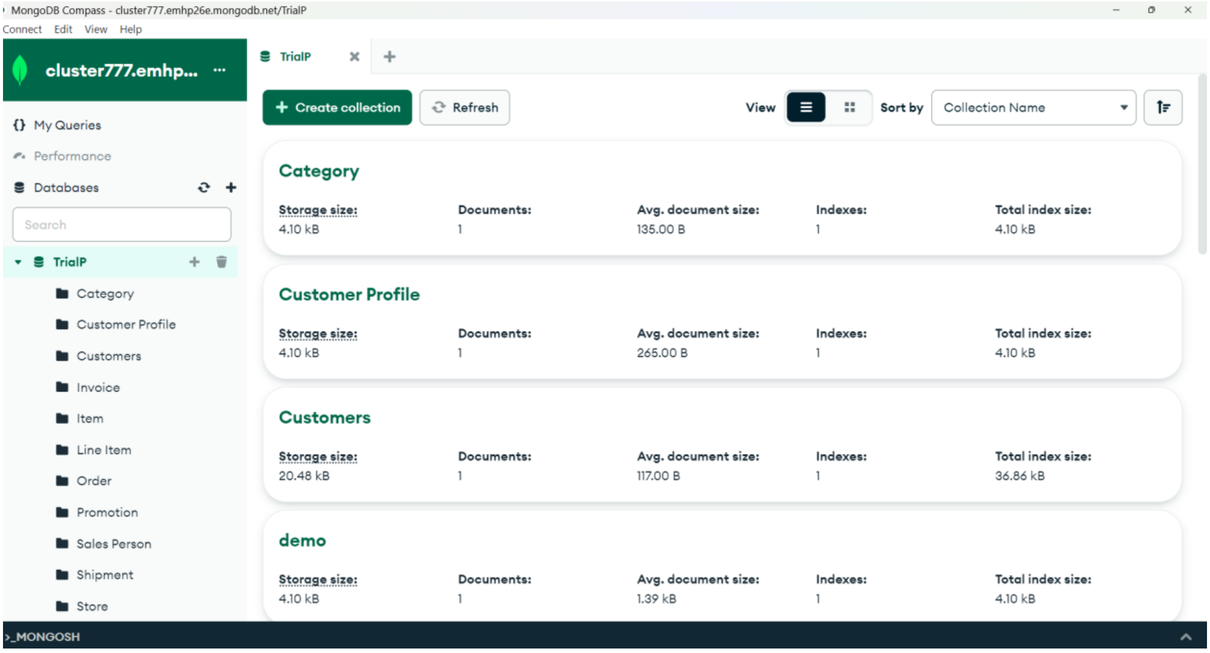


Figure 10:MongoDB Sample Database Documentation for Acme Inc.using Dummy Data

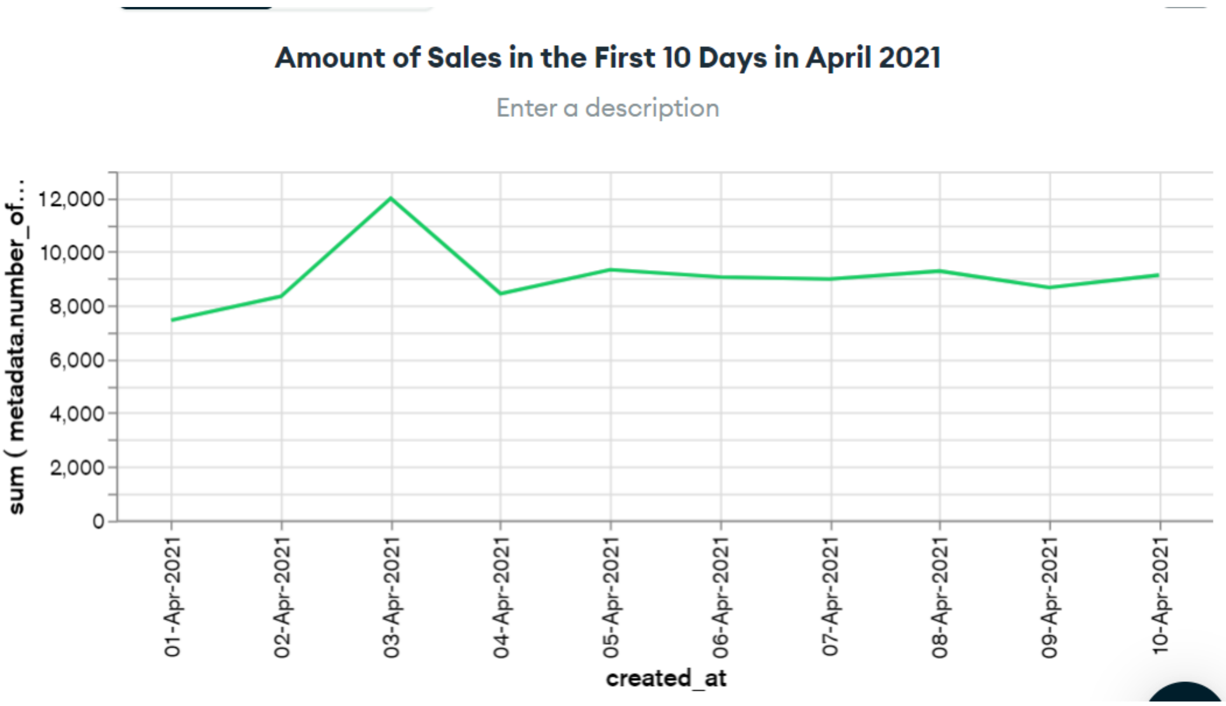


Figure 11:MongoDB Charts Sample Visualization (2)

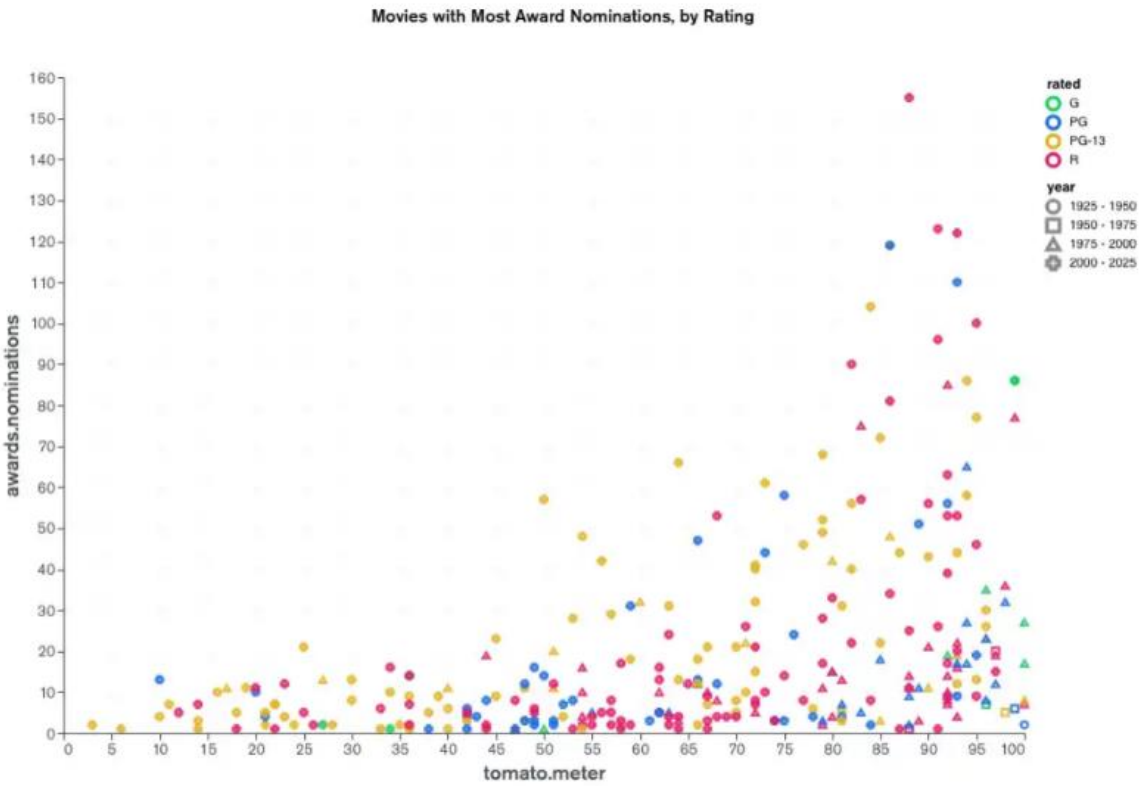


Figure 12:MongoDB Charts Sample Visualization (2)

Machine Learning and Predictive Models for Warehouse Management at Acme Inc.

Machine learning (ML) and predictive modeling present powerful tools for improving warehouse management efficiency. Acme Inc. faces challenges with inaccurate forecasting, delayed shipments, and damaged items. Implementing ML algorithms can address these issues by enhancing forecasting accuracy, optimizing inventory levels, and minimizing shipment delays (Abbott 2014).

Overview of Solution

By leveraging machine learning and predictive modeling, Acme Inc. can develop sophisticated algorithms to forecast demand accurately, identify potential shipment delays in advance, and predict factors leading to damaged items. These predictive models can be integrated into their existing IT systems for real-time decision-making.

Steps

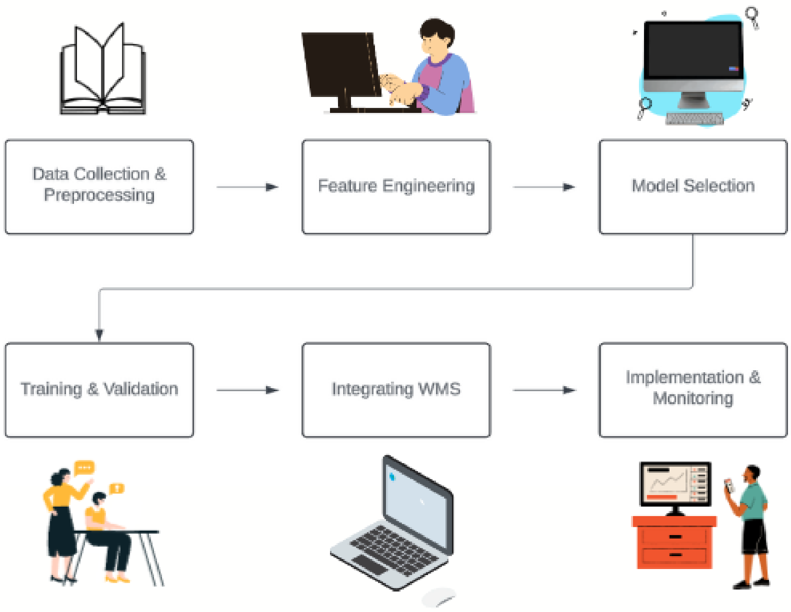


Figure 13:Summarized Steps to create ML & Predictive Models

Abbott (2014) highlighted the common steps to create and test the ML based model, presented below.

1. Data Collection and Preprocessing:
    - Gather historical sales data, shipment records, and information on damaged items.
    - Clean and preprocess the data to remove anomalies and ensure consistency.

2. Feature Engineering:
    - Identify relevant features such as SKU, time of year, promotions, and weather data (if applicable).
    - Create new features like seasonality indicators, lead times, and shipment routes.

3. Model Selection:
    - Choose appropriate ML models such as Random Forest, Gradient Boosting, or LSTM networks for time-series forecasting.
    - For shipment delay prediction, consider classification models like Logistic Regression or Support Vector Machines.
    - Use anomaly detection algorithms like Isolation Forest or Autoencoders to identify potential causes of damaged items.

4. Training and Validation:
    - Split the data into training and validation sets.
    - Train the selected models on historical data and fine-tune hyperparameters for optimal performance.
    - Validate the models using holdout validation or cross-validation techniques.

5. Integration with Warehouse Management System:
    - Develop APIs or connectors to integrate the trained models with Acme's warehouse management system.
    - Ensure real-time data flow between the models and the warehouse database.

6. Implementation and Monitoring:
    - Deploy the integrated ML models into the production environment.

- Continuously monitor model performance and retrain as new data becomes available.
- Implement alert systems for early detection of potential stock-outs, shipment delays, or factors leading to damaged items.

Benefits

A few benefits are expected to be gotten by Acme Inc. from utilizing the new system, i.e. improved forecasting accuracy, early shipment delay prediction, and damage prevention. The generated ML models will provide more accurate demand forecasts, reducing excess inventory and stock-outs. Predictive models will identify potential delays in shipments, allowing Acme to take proactive measures such as rerouting or expediting orders (Abbott 2014). In addition, Acme can implement preventive measures to reduce product damage, and loss, and improve customer satisfaction.

Conclusion

Machine learning and predictive modeling offer a comprehensive solution for Acme Inc.'s warehouse management challenges. By harnessing the power of data and advanced algorithms, Acme can optimize its inventory levels, streamline logistics, and improve overall operational efficiency.

Conclusion

In conclusion, Acme Inc. faces significant operational challenges that hinder its competitiveness and efficiency. Through strategic objectives aimed at optimising inventory management, enhancing promotional efficacy, and integrating IT infrastructure, Acme endeavours to construct a cohesive, data-centric infrastructure. By leveraging advanced analytics and data warehousing solutions, Acme seeks to refine its decision-making processes and improve operational agility. The proposed architectures, including Enterprise Data Warehousing, Data Marts, and cloud-based solutions, offer scalable and flexible frameworks to meet Acme's diverse business requirements. Additionally, the adoption of Neo4j-based solutions presents innovative opportunities for enhanced data analysis and visualization (Krishnamoorthy, 2014). However, successful implementation hinges on accurate sizing and estimation of hardware infrastructure, rigorous change management, and effective communication strategies to ensure organisational readiness and stakeholder engagement (Hassin, 2018). Through these concerted efforts, Acme Inc. aims to overcome its operational challenges and emerge as a more competitive player in the consumer electronics and durables sector.

Reflective Insights:

The process of delving into Acme Inc.'s operational challenges and crafting solutions has provided valuable insights into the complexities of modern business operations and the pivotal role of data-driven decision-making. Here are some reflective insights gleaned from this exercise:

- Integration Complexity: Acme's heterogeneous IT landscape underscores the challenges associated with integrating disparate systems and databases. Achieving seamless data flow and interoperability requires careful planning, collaboration across departments, and robust data governance frameworks (Nagabhushana, 2006).

- Strategic Alignment: It is critical that technology solutions and strategic business goals are in sync. Organisations such as Acme may optimise the return on their IT infrastructure expenditures by giving priority to initiatives that directly address operational pain points and support long-term growth plans (Hicks, 2000)

- Change Management Imperative: Transformative solutions like data warehousing require more than just technological know-how to implement. Stakeholder involvement, effective change management, and communication tactics are essential for breaking through opposition, building buy-in, and facilitating successful adoption throughout the organization (Hassin, 2018).

Continuous Iteration: Making decisions based on data is an iterative process that calls for dedication to constant development. To adjust their strategy in response to changing business requirements and technical breakthroughs, Acme must adopt a culture of experimentation, learning from both triumphs and mistakes (Fienberg, 2006).

By leveraging these insights and adopting a proactive, adaptive mindset, Acme Inc. can navigate the complexities of modern business environments and emerge as a more agile, data-driven organisation poised for sustainable success.

References:

- Abbott, D 2014, '*Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst'*, *Google Books*, John Wiley & Sons, viewed 5 April 2024, <https://books.google.com.au/books?hl=id&lr=&id=ZNVFAwAAQBAJ&oi=fnd&pg=PA1&dq=steps+to+build+predictive+analytics&ots=YYOD1eCtWD&sig=85oVMu3q-F76SmCgr9epu6roksA&redir_esc=y#v=onepage&q=steps%20to%20build%20predictive%20analytics&f=false>.

- Andiyappillai, N 2020, '*Standardization of System Integrated Solutions in Warehouse Management Systems (WMS) Implementations*', International Journal of Applied Information Systems, 12(35), 24-29, viewed 2 April 2024, <https:// doi: 10.5120/ijais2020451895>.

- Armstrong, R 1997, '*Data warehousing: dealing with the growing pains*', Proceedings 13th International Conference on Data Engineering, Data Engineering, 1997. Proceedings. 13th International Conference on, Data engineering, pp. 199–205, viewed 4 April 2024,<https://research.ebsco.com/linkprocessor/plink?id=7b51714a-0be8-31ed-b862-e86ca26e0322>.

- Arora, A & Gosain, A 2021, '*Mechanism for securing cloud based data warehouse schema*', International Journal of Information Technology: An Official Journal of Bharati Vidyapeeth's Institute of Computer Applications and Management, vol. 13, no. 1, pp. 171–184, viewed 3 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=f3c2bc4a-6fbd-3ce5-89f6-c821268c837c>.

- Bhatia, P 2019, '*Big Data and NoSQL in Data Mining and Data Warehousing: Principles and Practical Techniques*', pp. 442–466, viewed 3 April 2024, < https://doi.org/10.1017/9781108635592>.

- Blažić, G, Poščić, P & Jakšić, D 2017, '*Data warehouse architecture classification'*, 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2017, pp. 1491-1495, viewed 2 April 2024, <https:// doi: 10.23919/MIPRO.2017.7973657>

- Botoeva, E, Calvanese, D, Cogrel, B, & Xiao, G 2018, '*Expressivity and complexity of MongoDB queries*', In 21st International Conference on Database Theory (ICDT 2018), Schloss-Dagstuhl-Leibniz Zentrum für Informatik, viewed 2 April 2024, <https://doi.org/10.4230/LIPIcs.ICDT.2018.9>

- Campbell, A, Xiangbo Mao, Jian Pei & Al-Barakati, A 2017, 'Multidimensional benchmarking in data warehouses', *Intelligent Data Analysis*, vol. 21, no. 4, pp. 781–801, viewed 4 April 2024, https://research.ebsco.com/linkprocessor/plink?id=d6aea633-ea34-398b-8a6d-0c7d7286ffba

- Dias, MM, Tait, TC, Menolli, ALA & Pacheco, RCS 2008, '*Data Warehouse Architecture through Viewpoint of Information System Architecture*', IEEE Xplore, pp. 7–12, viewed 2 April 2024, <https://ieeexplore.ieee.org/document/5172591>.

- Edward, SG & Sabharwal, N 2015, *Practical MongoDB: Architecting, Developing, and Administering MongoDB*, *Google Books*, Apress, viewed 5 April 2024, <https://books.google.com.au/books?hl=id&lr=&id=7iI3CwAAQBAJ&oi=fnd&pg=PR6&dq=creating+interface+in+mongoDB&ots=cNtmSTXKMp&sig=fs68c215ewRiwswXWMzFDfanPlw&redir_esc=y#v=onepage&q=creating%20interface%20in%20mongoDB&f=false>.

- Erraji, A, Maizate, A, & Ouzzif, M 2022, '*New etl process for a smart approach of data migration from relational system to mongodb system*', In International Conference on Digital Technologies and Applications (pp. 131-140), Cham: Springer International Publishing, viewed 2 April 2024, <https://doi.org/10.1007/978-3-031-01942-5_13>.

- Fienberg, SE 2006, '*Privacy and Confidentiality in an e-Commerce World: Data Mining, Data Warehousing, Matching and Disclosure Limitation*', Statistical Science, vol. 21, no. 2, pp. 143–154, viewed 4 April 2024, https://research.ebsco.com/linkprocessor/plink?id=7f7b53fd-8621-35cb-930a-c462c925c3b0

- Gupta, A, Sahayadhas, A & Gupta, V 2020, '*Proposed Techniques to Design Speed Efficient Data Warehouse Architecture for Fastening Knowledge Discovery Process*', 2020 IEEE Third International Conference on Artificial Intelligence and Knowledge

Engineering (AIKE), Laguna Hills, CA, USA, 2020, pp. 200-201, viewed 1 April 2024, <https:// doi: 10.1109/AIKE48582.2020.00039>

- Hassin, A 2008, '*The importance of data warehousing for community development organisations: the impact on Iraqi victims of the explosive remnants of war*', viewed 7 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=b5cc1e8b-75fc-3724-ad07-7c093d085a10>.

- Herden, O 2020, '*Architectural Patterns for Integrating Data Lakes into Data Warehouse Architectures'*, Lecture Notes in Computer Science (), vol 12581, viewed 3 April 2024, <https://doi.org/10.1007/978-3-030-66665-1_2>

- Hicks, D & Nivin, S 2000, '*Beyond Globalization: Localized Returns to IT Infrastructure Investments'*, Regional Studies, vol. 34, no. 2, p. 115, viewed 6 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=8b82a385-8b2f-3874-9f37-6e278314f2bc>.

- Inmon W.H & Daniel Linstedt 2014, '*Data Architecture: A Primer for the Data Scientist : Big Data*', Data Warehouse and Data Vault, Morgan Kaufmann, Amsterdam, viewed 2 April 2024, https://research.ebsco.com/linkprocessor/plink?id=2fcfecb6-ab32-360a-9413-98011425d8f8

- Kemper, C 2015, '*Beginning Neo4j'*, viewed 3 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=adebf8b1-67f6-356d-9cb7-2fc0d93c67ba>.

- Kimelfeld, B & Amsterdamer Y (eds), Dagstuhl Research Online Publication Server, vol. 98, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, pp. 9:1–9:23, viewed 2 April 2024, <https://drops.dagstuhl.de/opus/volltexte/2018/8607/>.

- Kondylakis, H, Spanakis, E, G, Sfakianakis, S, Sakkalis, V, Tsiknakis, M, Marias, K, & Dong, F 2015, "*Digital patient: Personalized and translational data management through the MyHealthAvatar EU project*", IEEE, pp. 1397-1400, viewed 8 April 2024, <https://doi.org/10.1109/EMBC.2015.7318630>.

- Krishnamoorthy 2014, *Data Warehousing and Multi-Dimensional Data Modeling*, SAGE Publications, viewed 8 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=6427d269-be45-3ee0-8a23-22bcbb99e264>.
- Mourya, SK 2012, '*Data Mining and Data Warehousing'*, viewed 3 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=a76a2a4d-4cd4-37ba-a2fa-7fb54a2623eb>.

- Nagabhushana, S 2006, '*Data warehousing : OLAP and data mining'*, viewed 4 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=c735c498-b51e-38e1-a5c8-daa4b1651564>

- Prabhu, S 2007, '*Data mining and warehousing'*, viewed 3 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=baa894dd-d9c8-33ba-a3f6-3d0397eba735>.

- Qian, Z & Li-jun, S 2009, '*The Architecture and Design Strategy for Data Warehouse of Highway Management*,' Second International Conference on Intelligent Computation Technology and Automation, Changsha, China, 2009, pp. 459-462, viewed 2 April 2024, <https:// doi: 10.1109/ICICTA.2009.826>.

- Sun, L, Hu, M, Ren, K & Ren, M 2013, '*Present Situation and Prospect of Data Warehouse Architecture under the Background of Big Data'*, viewed 4 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=601201d2-53de-3d9e-a2a3-158582f517b9>.

- Swoyer, S 2021, '*Automating the Modern Data Warehouse',* viewed 3 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=b0416fd9-28dd-3ff9-9537-5392628b0f0e>.

- Vukotic, A 2015, '*Neo4j in action'*, viewed 3 April 2024, <https://research.ebsco.com/linkprocessor/plink?id=a4c975e3-bdc2-303f-b3b2-cf4254b0a83f>.