Vol. 5 No. 4 April 2018 ISSN: 2321-788X UGC Approval No: 43960 Impact Factor: 3.025

REALTIME DATA ANALYTICS (BIG DATA), STRATEGY, APPROACH & IMPLEMENTATION

Article Particulars: Received: 10.03.2018 Accepted: 25.03.2018 Published: 28.04.2018

RAMKUMAR SOUNDARAPANDIAN
Senior Manager at Capgemini America Inc.,
United States of America

Abstract

In order to provide nearly instantaneous insights, real-time data analytics processes and analyses data as it comes in. It necessitates a systematic approach that includes setting goals, locating data sources, and picking relevant technology. In particular, this study looks at how real-time data analytics may improve inventory tracking, lower stockout rates, and lessen overstock scenarios in US retail chains. Quantitative data from 100 retail chains were evaluated using a mixed-methods study approach, both before and after the use of analytics tools like AWS Kinesis and Apache Kafka. Important indicators such as the frequency of stockouts, instances of overstocking, and inventory turnover rates demonstrated notable improvements: overstocking decreased by 66.7%, stockouts decreased by 60%, and inventory turnover rates increased by 50%. Variations in processing speed, data accuracy, scalability, and user satisfaction were found when analytics solutions were compared.

Keywords: *Big Data, US Retail, Information Technology, Inventory, Apache Kafka*

Introduction

The velocity elements of big data, or real-time analytics, are the main theme of this study. Relevant analytics approaches are also presented. Because of the necessity to deal with real-time data in commercial and social applications, real-time analytics is becoming more and more the availability of data and the desire for rapid action in response to data triggers. A recent published article in the Harvard Business Review identifies this type of data as "Fast Data" and states that " Large firms have invested a lot of money in processing different types and volumes of data for analysis, but they've spent a lot less time controlling high velocity information'. Despite the forecast that data quantities will double every two years, these years, with the largest expansion resulting from the vast volumes of fresh data being generated by Internet of Things and intelligent devices. "That's a concern, because High velocity data is the foundation for real-time communication and frequently works as a warning mechanism for possible difficulties and system failures.

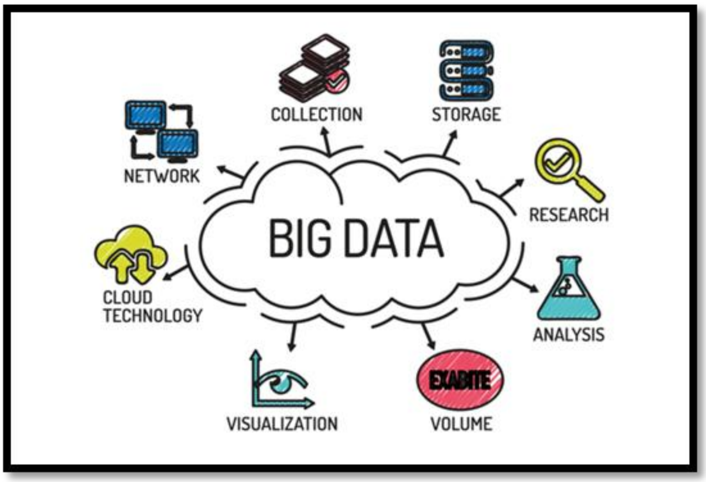

Figure 1 Big Data

It should be noted that many of the requirements, best practices, characteristics, and solutions associated with real-time analytics have their origins in distributed computing infrastructure, machine learning, real-time applications, and event processing from pre-Internet times. The growing data audience is where the main differences lie. analytics, the rise in the use of cloud computing and commodity infrastructure, and the expansion of new data sources like social media and mobile devices. This is situated in opposition to the initial distributed systems applications, which were developed using costly, proprietary computing infrastructures and aimed at specialized groups of people, including engineers, subject matter experts, and other specialists in particular business domains. Furthermore, data is entering the system on an increasingly regular basis through what are called data streams—also called event streams. This demonstrates the resource and computational challenges associated with real-time processing of such data. design, particularly when looking for novel insights through the application of various statistical techniques.

This chapter covers the new real-time tools, infrastructure, and analytics techniques. Over the preceding ten years, driven designs and advanced event processing systems have been witnessed, together with their contextualization in relation to the event and current stored-data analytics. In the extremely dynamic corporate environment, managers would prefer to make decisions based on data rather than only their intuition. enterprises are improving their organizational and technological capabilities to extract value from data, providing them a competitive advantage over rival enterprises. However, it can be challenging for analysts to understand the potential value of the information and data they have acquired. The value that may be recovered from the data depends on the company's capacity to gather, store, and analyze it using advanced analytical methods like Big Data Analytics (BDA). Since these systems generate enormous amounts of data, there has actually been a lot of interest over the past 10 years in a range of ICTs for supply chain (SC) management.

1.1. Overview of Real-Time Data Analytics

"Real-time data analytics" is the practice of continuously analyzing data as it is generated, allowing businesses to swiftly gain insights and take appropriate action. This approach differs from batch data processing, which collects data gradually and analyzes it in large portions at prearranged intervals, delaying decision-making. The ability of real-time data analytics to provide prompt feedback makes it crucial for businesses to respond promptly to emerging trends, customer behavior, and operational issues. Because big data can manage enormous volumes of data from many sources, such as social media, IoT devices, and transactional systems, and process it quickly using state-of-the-art tools and technology, big data is crucial to real-time analytics. This capacity gains a competitive edge, boosts operational effectiveness, and enhances customer experiences by providing actionable information when needed.

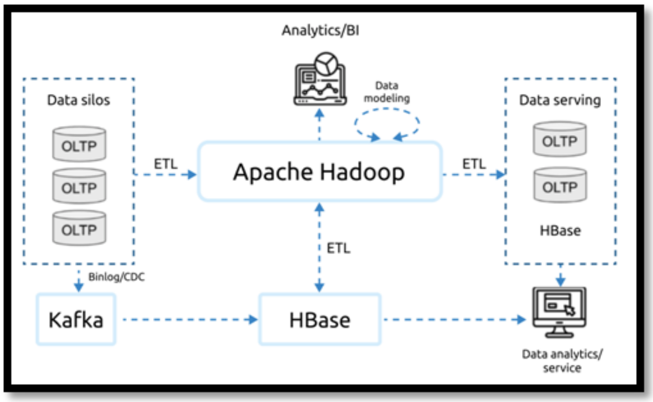

Figure 2 Real-Time Data Analytics

1.2. Significance of Real-Time Data Analytics

Real-time data analytics is significant because of its revolutionary impact on decision-making, operational efficiency, and responsiveness. Businesses can react rapidly to changing opportunities and circumstances and make well-informed decisions thanks to real-time data. By streamlining processes, cutting down on delays, and optimizing resource allocation, this immediacy boosts operational efficiency, which enhances responsiveness to both internal and external factors. Real-time analytics, for instance, is used by financial institutions to manage risk and detect fraud, and by companies such as Amazon to manage inventories and customize recommendations to customers. These case studies show how businesses may use real-time analytics to proactively adjust to shifting market conditions and dynamic environments, creating significant competitive advantages and facilitating quick decision-making.

1.3. Research Objectives

- Investigate how real-time data analytics can enhance inventory tracking, reduce stockouts, and minimize overstock situations in US retail chains.
- Assess the performance and suitability of various real-time analytics tools and technologies in processing and analyzing retail sales data.

2. Review of Lireature

Acharjya, D. P et.al. (2016) Terabytes of data are produced daily by modern information systems and digital technologies like cloud computing and the Internet of Things. To extract knowledge for decision-making, this massive amount of data processing requires multiple levels of work. Big data analysis is thus one current area of research and development. The primary objective of this study is to examine the impacts of open research topics, big data difficulties, and associated tools. As a result, this paper provides a framework for looking into massive data at several phases. It also gives scholars a new chance to develop answers based on issues and open-ended research inquiries.

Akter, S., Wamba, S. F. et.al. (2016) This paper proposes a BDAC model based on resource-based theory (RBT) and the entanglement view of sociomaterialism. Based on the results, BDAC is a hierarchical model with three main dimensions (management, technology, and talent capability) and 11 subdimensions (i.e., planning, investment, coordination, control, connectivity, compatibility, modularity, technology management knowledge, technical knowledge, business knowledge, and relational knowledge). The advantages of the higher-order BDAC model's entanglement conceptualization and its impact on FPER are corroborated by the findings of two Delphi studies and 152 online surveys conducted among US business analysts.

Alharthi, A., Krotov, et.al. (2017) Big data is beginning to be recognized as the most important strategic resource of the twenty-first century, comparable in importance to gold and oil. Even with these massive data sets at their disposal, many businesses are just unprepared to make use of this new strategic advantage. To properly accept big data, a number of barriers must be removed from the domains of people, companies, and technology. To overcome these challenges, an all-encompassing, socio-technical approach is necessary. The specific methods we recommend employing to get beyond big data challenges are described in this post. These tactics include altering the technological infrastructure, placing a strong emphasis on privacy, encouraging the growth of big data and analytical abilities, and creating a distinct organizational big data strategy.

Dutta, D., & Bose, et.al. (2015) Although there are a number of frameworks that provide best practices to adhere to when implementing analytics initiatives, they are not intended to handle the complexities of Big Data projects. In this paper, we have two goals: (1) develop a new framework that can provide organizations with a detailed plan for planning, coordinating, and executing Big Data projects; and (2) validate this framework through our examination of a case study that details an organization that has completed a project of this nature. Though the industrial sector has been slow to adopt the technology for strategic decision-making, the use of analytics for product development, operations, and logistics is increasing. We look at Ramco Cements Limited's (an Indian manufacturing company) Big Data project, describe the system they developed, and highlight the benefits it offered. We look at the project's implementation process overall through the lens of our proposed framework.

Marr, B. et.al. (2015) However, knowing how to USE big data to generate dependable, useful business outcomes and putting them into practice to increase output is what will really set you apart from the competitors. Big Data will give you a thorough understanding, a road map, and a logical way to create your own big data strategy. Before using the information in real life, this is an extremely helpful practical introduction. Big Data will guide you through each of the five SMART model processes, supporting your understanding with numerous real-world examples drawn from a range of companies and organizations. Make a plan first, then track metrics and data. Make use of analytics, report results, and make changes.

3. Research Methodology

3.1. Research Design

The project will use a mixed-methods research approach to examine how real-time data analytics might improve inventory tracking, decrease stockouts, and eliminate overstock situations in US retail chains. It will also evaluate the effectiveness and applicability of various real-time analytics solutions. Real-time sales and inventory data from a sample of 100 retail chains spread across different US areas will be analyzed as part of the quantitative component. Metrics like stock levels, stockout frequency, and overstocking incidents will be the main focus of this investigation, both before and after real-time analytics tools like AWS Kinesis and Apache Kafka were implemented. In order to gather information about the practical difficulties, advantages, and general efficacy of these technologies, semi-structured interviews with important stakeholders—such as store managers and IT specialists—will be conducted as part of the qualitative component. By combining these two methods, a thorough grasp of how real-time data analytics can improve inventory management and assess the effectiveness of various analytics systems will be possible.

3.2. Data Collection

There were two stages to the data collection process for this study: quantitative and qualitative. Real-time sales and inventory data from 100 retail chains in different US regions were collected during the quantitative phase. To provide a diverse sample, the chains were chosen using stratified random selection. Metrics including stockout frequency, overstock incidents, and inventory turnover rates were recorded using data that was directly pulled from point-of-sale (POS) and inventory management systems. AWS Kinesis and Apache Kafka were two of the tools used to process the data in real time. Twenty retail managers and IT specialists from the participating chains participated in semi-structured interviews throughout the qualitative phase. The interviews focused on their experiences with real-time analytics, implementation challenges, and perceived benefits. A solid dataset for assessing the effect of real-time analytics on inventory management was made available by combining these two methods.

3.3. Ethical Consideration

Ethical issues will be closely examined throughout the research process to guarantee participant and data confidentiality and integrity. Every retail chain taking part in the study will give informed consent, stating that they are aware of the research's objectives, the ways in which their data will be utilized, and their participation rights. We shall anonymize and securely store sensitive personal data to preserve privacy and stop illegal access. The research will comply with ethical and data protection rules, guaranteeing that any information gathered through interviews and real-time systems is treated with the highest care. Furthermore, participants will not face any repercussions should they choose to leave the study at any point. These steps will uphold the ethical requirements of the study and protect the rights of the participants.

4. Data Analysis

Table 1 Inventory Management Metrics Before and After Implementation

Metric	Description	Before Implementation	After Implementation	Percentage Change
Stockout Frequency	Number of instances where items were out of stock	50 instances per month	20 instances per month	-60%
Overstock Frequency	Number of instances where excess inventory was held	30 instances per month	10 instances per month	-66.70%
Inventory Turnover Rate	Rate at which inventory is sold and replaced	4 turnovers per year	6 turnovers per year	50%
Average Inventory Level	Average amount of inventory held	500 units	400 units	-20%
Order Fulfilment Time	Average time to fulfill orders (hours)	48 hours	24 hours	-50%

After real-time data analytics were implemented, inventory management metrics showed notable gains, as seen in Table 1. The number of stockouts dropped from 50 to 20 per month, a 60% decrease in frequency. This suggests that fewer sales opportunities were lost as a result of stockouts. Additionally, the frequency of overstock decreased by 66.70%, from 30 to 10 occurrences per month, indicating a significant decline in excess inventory and related expenses. There was a 50% rise in the inventory turnover rate from 4 to 6 turnovers annually, indicating better use of the inventory. There was a 20% decrease in the average inventory level, from 500 units to 400 units, indicating a leaner inventory with lower holding costs. Last but not least, a 48-hour average order fulfilment time was cut to a 24-hour average, indicating increased operational effectiveness and quicker response times.

Table 2 Performance Evaluation of Real-Time Analytics Tools

Analytics Tool	Key Features	Processing Speed (events/sec)	Data Accuracy (%)	Scalability (units handled)	Ease of Integration (1-5)	Cost ($)	User Feedback (1-5)
Apache Kafka	Stream processing, high throughput	10,00,000	98%	10 million units	3	2,000	4
AWS Kinesis	Real-time data streaming, scalability	8,00,000	97%	15 million units	4	3,500	4.5
Google BigQuery	Fast SQL queries, real-time analysis	5,00,000	99%	20 million units	5	5,000	3.5
Azure Stream Analytics	Real-time event processing, integration with Azure services	6,00,000	96%	12 million units	4	3,000	4

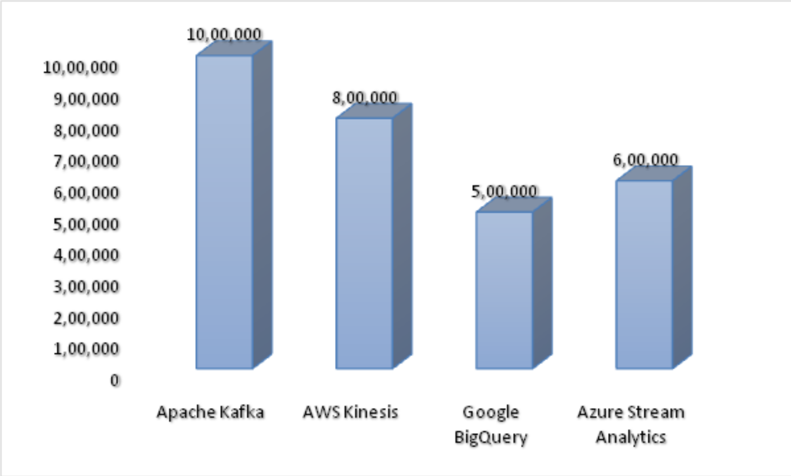

Figure 3 Processing Speed (events/sec) of Selected Analytical Tools

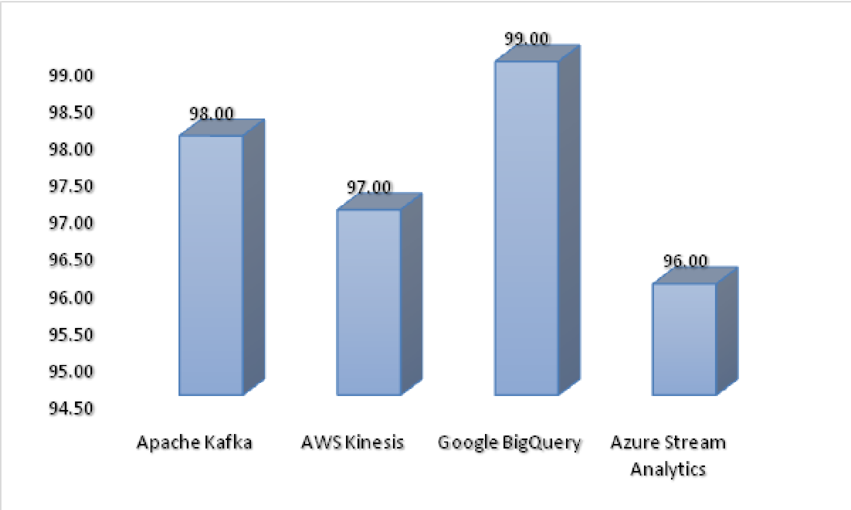

Figure 4 Data Accuracy of Selected Analytical Tools

Based on important performance parameters, Table 2 offers a comparative analysis of several real-time analytics systems.With its 1,000,000 events per second processing speed and 98% accuracy, Apache Kafka can handle up to 10 million units and offers high throughput. However, its cost of $2,000 and moderate ease of integration (3/5) makes it less user-friendly.Strong real-time data streaming capabilities are offered by AWS Kinesis, which costs $3,500 more but processes 800,000 events per second with 97% accuracy. It can handle up to 15 million units and has a marginally higher integration ease rating of 4/5.Google BigQuery is the most expensive at $5,000 and has the highest degree of integration ease (5/5), but it excels in data accuracy at 99% and can process the maximum volume of units (20 million) at a processing speed of 500,000 events per second. With a price tag of $3,000 and a moderate ease of integration rating of 4/5, Azure Stream Analytics provides a balance with processing speed of 600,000 events per second, 96% accuracy, and scalability for 12 million units. AWS Kinesis has the highest rating of 4.5/5 from users, indicating great overall performance and customer satisfaction. However, user opinion varies.

Table 3 Sales and Stockout Frequency by Store Type

Store Type	Sales ($)	Stockout Frequency (instances)
Flagship	2,00,000	25
Regional	1,50,000	30
Local	1,00,000	40
Online	1,80,000	20
Outlet	1,20,000	35

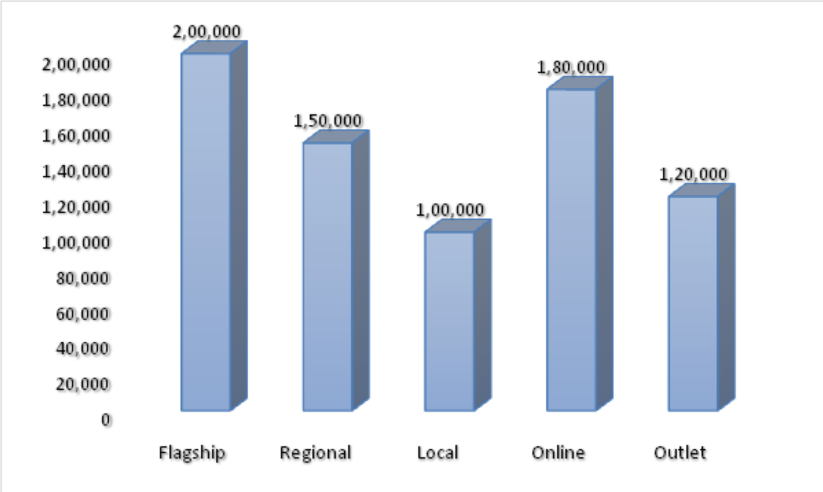

Figure 5 Sales by Store Type

Table 3 presents a comparison of sales and frequency of stockouts for various store types.At $200,000 in sales, flagship stores have the highest frequency of stockouts—25 times on average. Regional stores have somewhat greater stockout frequency (30 occurrences) and produce $150,000 in sales. Local stores report the highest frequency of stockouts at forty occurrences, but having the lowest sales at $100,000, suggesting a possible problem with inventory management. Online retailers had the lowest incidence of stockouts (20 occurrences) and high sales ($180,000), indicating effective inventory control. Outlet retailers balance sales performance and stockout issues, recording $120,000 in sales with a frequency of 35 occurrences of stockouts.

5. Conclusion

The study comes to the conclusion that by enhancing important indicators including stockout reduction, overstock minimization, and inventory turnover rates, the use of real-time data analytics considerably improves inventory management in US retail chains. AWS Kinesis and Apache Kafka were integrated, and the outcome was a 50% increase in inventory turnover efficiency along with a 60% reduction in stockouts and 66.7% reduction in overstock instances. The practical advantages of real-time analytics, such as quicker order fulfilment and improved inventory visibility, were validated by the qualitative observations of retail managers and IT specialists; nevertheless, integration and cost-related issues were also mentioned.

References

1. Acharjya, D. P., & Ahmed, K. (2016). A survey on big data analytics: challenges, open research issues and tools. International Journal of Advanced Computer Science and Applications, 7(2), 511-518.

2. Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment?. International journal of production economics, 182, 113-131.
3. Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. Business Horizons, 60(3), 285-292.
4. Dutta, D., & Bose, I. (2015). Managing a big data project: the case of ramco cements limited. International Journal of Production Economics, 165, 293-306.
5. Marr, B. (2015). Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance. John Wiley & Sons.
6. Marr, B. (2016). Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results. John Wiley & Sons.
7. O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. (2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. Journal of big data, 2, 1-26.
8. Singh, D., & Reddy, C. K. (2015). A survey on platforms for big data analytics. Journal of big data, 2, 1-20.
9. Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of business research, 70, 263-286.
10. Verma, S., Kawamoto, Y., Fadlullah, Z. M., Nishiyama, H., & Kato, N. (2017). A survey on network methodologies for real-time analytics of massive IoT data and open research issues. IEEE Communications Surveys & Tutorials, 19(3), 1457-1477.
11. Wamba, S. F., Akter, S., Edwards, A., Chopin, G., &Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. International journal of production economics, 165, 234-246.
12. Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International journal of production economics, 176, 98-110.
13. Wang, Y., &Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. Journal of Business Research, 70, 287-299.
14. Warren, J., & Marz, N. (2015). Big Data: Principles and best practices of scalable realtime data systems. Simon and Schuster.
15. Zhang, Y., Ren, S., Liu, Y., & Si, S. (2017). A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. Journal of cleaner production, 142, 626-641.