World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Check for updates

Big data and distributed computing: Transforming marketing through real-time customer data platforms

Gowri Shankar Ravindran *

Anna University, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(03), 1144-1150

Publication history: Received on 25 April 2025; revised on 03 June 2025; accepted on 06 June 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.3.0962

Abstract

Big data and distributed computing technologies transform marketing through real-time customer data platforms (CDPs). Platforms unify first-party data from diverse digital touchpoints into comprehensive customer profiles that can be activated instantly across channels. Technical foundations of CDPs include distributed computing frameworks, stream processing technologies, and identity resolution capabilities that support real-time data management at scale. Sophisticated customer segmentation and personalization technologies create relevant experiences delivered at the moment of interaction. Emerging trends highlight artificial intelligence integration, privacy-preserving computation, and engagement-focused marketing approaches while balancing personalization expectations and privacy considerations.

Keywords: Customer Data Platforms; Distributed Computing; Real-Time Personalization; Marketing Analytics; Customer Experience Management

1. Introduction the convergence of big data and marketing

The digital transformation of consumer behavior has generated unprecedented volumes of interaction data, creating both challenges and opportunities for marketing practitioners. Traditional data management approaches have proven inadequate for capturing, processing, and activating the velocity, volume, and variety of information generated through modern customer journeys. In the advent of AI and high-scale distributed vector stores, semantic understanding of customer profiles becomes a critical asset for marketing professionals. Social media platforms alone generate enormous quantities of new data daily, with billions of social media users worldwide creating, sharing, and engaging with content at rapidly accelerating rates. Personalized marketing at scale and segment-based communication are complementary in nature. The marketing significance of this data explosion becomes apparent when considering that a substantial majority of consumers discover new products on social platforms, and most make purchase decisions influenced by social content. These patterns underscore the critical need for advanced analytics capabilities as highlighted in comprehensive research examining the power of social media analytics [1].

Customer Data Platforms (CDPs) represent a significant evolution in marketing technology, designed specifically to unify customer data from multiple sources into coherent, persistent profiles that can be accessed and activated by other systems. Salesforce's comprehensive "State of the Connected Customer" research reveals that most customers now expect companies to understand their unique needs and expectations across all interaction channels. This expectation has driven organizations to adopt more sophisticated data management solutions, with many business leaders indicating that their digital transformation initiatives are directly focused on creating more connected customer experiences. The study further indicates that a majority of customers now expect personalized interactions, yet a substantial expectation gap exists, with only a minority of companies currently delivering personalization that

* Corresponding author: Gowri Shankar Ravindran.

customers consider excellent. This disconnect has fueled the rapid adoption of CDP technology, with the global CDP market growing at a significant rate annually [2]. In the advent of AI and high-scale distributed vector stores, semantic understanding of customer profiles becomes a critical asset of a marketing professional. Unlike previous generations of marketing databases, CDPs leverage distributed computing architectures to provide both comprehensive customer views and real-time processing capabilities. Personalized marketing at scale and segment-based communication are complementary in nature. This technological approach has fundamentally altered the possibilities for personalized marketing at scale, shifting the industry from segment-based communications toward individually tailored experiences delivered at the moment of customer interaction.

This article examines how the integration of big data technologies and distributed computing frameworks within CDPs is transforming marketing practices across industries, with particular focus on the mechanisms enabling real-time customer engagement and the resulting impact on marketing effectiveness.

2. Technological Foundations: Big Data Architecture and Distributed Computing

Real-time CDPs rely on sophisticated technological infrastructures that combine multiple big data components. At their core, batch-based distributed computing frameworks such as Apache Hadoop, Spark provide foundational data processing capabilities. Real-time distributed computing frameworks leveraging Apache Kafka with Apache Storm and Flink enable stream processing and event-driven architectures. According to detailed technical analysis, Apache Spark has emerged as a particularly powerful foundation for CDP implementations due to its ability to process data many times faster than traditional Hadoop MapReduce operations when running in memory, and considerably faster when running on disk.

The architecture of modern CDPs typically incorporates several key components working in concert. Distributed file systems provide the foundation for storing raw customer data at scale, with systems engineered to ingest both structured and unstructured data from disparate sources while maintaining sub-second access latencies. Stream processing technologies handle real-time data flows, processing interaction events as they occur rather than in overnight batches. Specialized graph databases optimize identity resolution and profile management, maintaining billions of entity relationships with query response times measured in milliseconds. Comprehensive application programming interfaces (APIs) support bidirectional data exchange, with enterprise-grade CDPs processing numerous API calls per second during peak operations. Advanced machine learning frameworks enable predictive analytics and automated segmentation, with models continuously retraining on newly acquired customer data. According to economic impact studies, organizations implementing modern CDP architectures experience substantially faster data processing capabilities compared to traditional database systems, enabling near-instantaneous customer data activation [3].

Table 1 CDP Component Architecture [3]

Component	Function	Technologies	Benefits
Data Collection	Capture customer data	Tags, SDKs, APIs	Complete customer view
Identity Resolution	Connect identifiers	Matching algorithms	Unified cross-channel profiles
Data Storage	Maintain profiles	Distributed systems	Scalable repository
Processing Engine	Segment customers	Stream processing	Real-time segmentation
Decision Engine	Determine next actions	ML models, rules	Personalized experiences
Activation Layer	Deliver experiences	APIs, webhooks	Cross-channel coordination
Governance	Ensure compliance	Consent management	Privacy and data quality

Enterprise-grade CDPs now routinely manage customer databases with enormous numbers of individual profiles, with the most sophisticated implementations containing hundreds of distinct attributes per profile. The total economic impact analysis of CDP implementations reveals that a significant majority of organizations report substantial improvements in data processing efficiency, with customer segmentation operations that previously required several hours now completing in just minutes. These technological advances create an ecosystem capable of continuously ingesting data from websites, mobile applications, CRM systems, point-of-sale terminals, and other customer touchpoints. Forrester's comprehensive analysis indicates that organizations implementing CDP technology achieve a

substantial ROI over multiple years, with the initial investment paying for itself rapidly for enterprise deployments [3]. The distributed nature of these systems enables organizations to maintain performance even as data volumes grow exponentially, providing a technical foundation for real-time marketing operations that would be impossible with centralized computing approaches.

3. Customer Profile Development: From Data Collection to Unified Identity

The primary function of CDPs is to construct comprehensive, 360-degree profiles of individual customers by unifying data from disparate sources. This process begins with the collection of first-party data—information gathered directly from customer interactions with an organization's owned channels. Sometimes it is complemented by second- and third-party data. Salesforce's extensive consumer research indicates that the average enterprise now tracks hundreds of discrete customer data points across digital properties, with the vast majority of organizations increasing their data collection efforts over the past few years.

Creating unified customer profiles presents significant technical challenges, particularly in resolving identities across devices and channels. Distributed computing frameworks enable CDPs to employ sophisticated identity resolution algorithms that can process billions of identity markers simultaneously. Social media analytics research reveals that the complexity of this challenge has grown exponentially, with the average consumer now using multiple connected devices and maintaining several active online accounts. Identity resolution must therefore reconcile disparate signals including authenticated identifiers like account logins and email addresses, persistent cookies and device IDs that may change over time, probabilistic matching based on behavioral patterns, and cross-device graphing through statistical inference. According to technical analyses of advanced matching algorithms, these systems now achieve high accuracy rates in cross-channel identity resolution, representing a dramatic improvement over the match rates typical of legacy marketing databases [1].

Once unified, these comprehensive profiles contain rich information about customer preferences, behaviors, and interaction history. The Forrester Total Economic Impact study documents that customer profiles within modern CDPs contain hundreds of discrete data points spanning demographic, behavioral, transactional, and contextual dimensions. Organizations leveraging these unified profiles report significant improvements in marketing efficiency, substantial increases in conversion rates, and higher average transaction values. Perhaps most significantly, the study indicates major improvement in marketing team productivity through automated data unification, with data engineers and analysts spending considerably less time on manual data preparation tasks [3]. The distributed architecture of CDPs allows these profiles to be continuously updated as new data becomes available, maintaining a current and accurate view of each customer that can be instantly accessed when needed for marketing activation.

4. Real-Time Segmentation and Activation Capabilities

With unified customer profiles as a foundation, marketers leverage the computational power of distributed systems to segment their audience with unprecedented granularity. Traditional marketing segmentation typically relied on basic demographic or purchase history variables, often processed in overnight batch operations. In contrast, CDP-enabled segmentation can incorporate thousands of behavioral and contextual variables, creating highly specific audience groups based on complex combinations of attributes and actions. The comprehensive framework for digital marketing established in the International Journal of Research in Marketing identifies this shift toward dynamic segmentation as one of the most significant evolutions in modern marketing practice. The research highlights how the digital environment has fundamentally transformed customer journeys from linear progressions to complex, multichannel interactions that generate vast amounts of behavioral data. This transformation necessitates new approaches to segmentation that can account for the contextual richness of customer interactions across touchpoints. The study further examines how digital technologies have reduced search costs for consumers while simultaneously increasing marketing capabilities for firms, creating an environment where segmentation accuracy directly correlates with competitive advantage [5].

Distributed computing frameworks allow these sophisticated segments to be generated dynamically and in real-time, supporting a diverse array of marketing use cases. Website personalization based on browsing behavior has emerged as a widely implemented capability, with the research framework for understanding and managing the customer experience emphasizing the critical role of contextualized digital interactions. This comprehensive research on customer experience management establishes that personalization effectiveness depends heavily on the timeliness of implementation, with real-time adaptations demonstrating significantly higher engagement metrics compared to delayed personalization. The framework further explores how triggered email campaigns responding to specific

customer actions represent a critical application of real-time segmentation. The research establishes that experience management across the customer journey requires both proactive and reactive interventions, with real-time triggered communications serving as a primary mechanism for addressing emerging customer needs at critical decision points. Mobile app notifications tailored to user location and context represent another application area described within the experience management framework, with the research emphasizing the heightened relevancy of contextually-aware messaging. Personalized product recommendations updated with each interaction demonstrate the continuous nature of modern customer experience management, with the research framework emphasizing the need for organizations to evolve from episodic to continuous engagement models [6].

Table 2 CDP Use Cases [6]

Use Case	Challenge	Solution	Key Metrics
Website Personalization	Static experiences	Real-time content adaptation	Conversion rate
Email Marketing	Generic campaigns	Triggered contextual messaging	Click-through rate
Advertising	Wasted ad spend	Precision targeting	Acquisition cost
Customer Service	Fragmented customer view	Unified profiles for agents	Resolution rate
Cross-Channel	Disconnected messaging	Synchronized experiences	Cross-channel conversion
Recommendations	Generic products	Real-time personalization	Items per order
Churn Prevention	Reactive approach	Proactive identification	Retention rate

The ability to process and activate these segments in milliseconds rather than hours represents a fundamental shift in marketing capabilities. Research on marketing analytics for data-rich environments published in the Journal of Marketing examines the technological infrastructure required to support real-time marketing operations. The research establishes a comprehensive framework for understanding analytics applications across the marketing function, with particular emphasis on the transformational impact of real-time capabilities. The study examines how increases in computational power through distributed processing have enabled marketing organizations to transition from descriptive to prescriptive analytics, with real-time segmentation representing a particularly valuable application. The research further identifies specific technological requirements for effective marketing analytics implementations, including data integration architecture, analytical model deployment mechanisms, and decision automation systems. The framework establishes that analytics sophistication exists along a continuum, with real-time activation representing the most advanced implementation stage. The study specifically addresses how the proliferation of customer touchpoints has created unprecedented segmentation opportunities while simultaneously increasing technical complexity, necessitating distributed computing approaches that can process disparate data streams concurrently [7].

5. Impact on Marketing Effectiveness and Customer Experience

The implementation of real-time CDPs powered by big data and distributed computing technologies has demonstrated measurable impacts on marketing performance across multiple dimensions. The digital marketing framework established in the International Journal of Research in Marketing explores how these technological capabilities translate into tangible business outcomes.

The research examines the full spectrum of digital marketing applications, establishing that data integration and real-time activation capabilities represent foundational requirements for optimizing customer acquisition and retention metrics. The study further explores how digital marketing has evolved through distinct stages, from channel-specific tactics to integrated approaches enabled by comprehensive data platforms. The framework specifically addresses how real-time personalization capabilities enable organizations to deliver contextually relevant content throughout the customer journey, enhancing engagement at critical decision points. The research also examines how marketing performance measurement has evolved alongside technological capabilities, with advanced attribution models leveraging unified customer data to provide more accurate effectiveness metrics [5].

Table 3 Impact on Marketing Effectiveness [5]

Dimension	Traditional Approach	CDP-Enhanced Approach	Impact
Segmentation	Static demographic	Dynamic behavioral	Higher response rates
Personalization	Rule-based, limited	ML-driven, complex	Increased conversion
Campaigns	Channel-specific	Orchestrated omnichannel	Improved journey
Analytics	Siloed, delayed	Unified, real-time	Optimized spending
Experience	Fragmented	Consistent, personalized	Increased loyalty
Team Efficiency	Manual processes	Automated workflows	Reduced costs
Privacy	Third-party cookies	First-party data strategy	Regulatory compliance

The framework for understanding and managing the customer experience published in the Marketing Science Institute Working Paper Series provides comprehensive insight into how real-time capabilities translate into enhanced customer relationships. The research establishes a multidimensional framework for conceptualizing customer experience, encompassing cognitive, emotional, sensory, social, and physical dimensions. The study examines how technological capabilities enable organizations to address each experience dimension more effectively, with real-time data platforms serving as a critical enabler of experience orchestration across touchpoints. The framework specifically addresses how consistency across channels represents a significant challenge for experience management, with unified customer data platforms providing the foundation for coherent cross-channel engagement. The research further explores the relationship between customer experience quality and business outcomes, establishing that experience improvements correlate strongly with increased loyalty, share of wallet, and advocacy behaviors. The framework also examines how experience measurement methodologies have evolved alongside technological capabilities, with real-time satisfaction monitoring providing more actionable insights than traditional measurement approaches [6].

Beyond these quantitative benefits, the strategic value of real-time customer data activation lies in its ability to meet evolving consumer expectations. The research on marketing analytics for data-rich environments published in the Journal of Marketing provides a comprehensive examination of how analytical capabilities enable enhanced customer understanding. The study establishes a framework for leveraging analytics throughout the marketing function, with particular emphasis on applications that enhance customer centricity. The research specifically addresses how advanced analytics enable organizations to transition from segment-based to individual-level engagement, creating more relevant and valuable customer interactions. The framework further explores how analytics sophistication correlates with marketing effectiveness, establishing that organizations with mature analytical capabilities demonstrate superior performance across acquisition, conversion, growth, and retention metrics. The research also examines how privacy considerations intersect with analytical capabilities, establishing a balanced approach that leverages first-party data while respecting consumer privacy preferences [7].

6. Future Directions and Conclusion

As big data and distributed computing technologies continue to evolve, their application within marketing contexts will likely expand in several directions. The digital marketing framework established in the International Journal of Research in Marketing outlines a comprehensive research agenda addressing emerging technological capabilities and their potential marketing applications. The research specifically identifies several critical areas for future development, including enhanced integration between marketing and business systems, improved attribution methodologies leveraging unified customer data, and more sophisticated predictive capabilities enabled by machine learning algorithms.

The framework further explores how technological advancements are reshaping the relationship between brands and consumers, creating opportunities for more collaborative and co-creative engagement models. The research also examines how regulatory developments will influence data collection and utilization practices, with particular emphasis on privacy-preserving technologies that enable personalization while respecting consumer rights [5].

Table 4 Emerging Trends in CDP Technology [6]

Trend	Current State	Future Direction	Strategic Impact
AI Integration	Basic predictive models	Advanced deep learning	Proactive marketing
Privacy Computation	Centralized processing	Federated learning, edge computing	Compliance with personalization
Edge Computing	Cloud-based	Processing at data source	Reduced latency
Architecture	Monolithic platforms	Modular, API-first components	Flexible ecosystem
Decision Systems	Rules with oversight	Self-optimizing algorithms	Scale and efficiency
Data Integration	Marketing-focused	Enterprise-wide unification	Complete experience management
XR Integration	Limited	Physical-digital-virtual integration	Metaverse personalization

The theory of customer engagement marketing published in the Journal of the Academy of Marketing Science provides valuable insight into how future technological developments may enhance engagement capabilities. The research establishes a comprehensive theoretical framework for conceptualizing customer engagement, encompassing behavioral, emotional, and cognitive dimensions. The study examines how engagement initiatives create reciprocal value for both customers and organizations, with technology serving as a critical enabler of scalable engagement strategies. The framework specifically addresses how engagement marketing represents an evolution beyond traditional relationship marketing approaches, with emphasis on fostering active customer participation rather than passive consumption. The research further explores how engagement manifests across different marketing contexts, with digital environments enabling particularly rich engagement opportunities. The theoretical framework also addresses measurement challenges associated with engagement marketing, establishing methodologies for quantifying engagement intensity and its relationship to business outcomes. The research concludes by examining how engagement marketing approaches may evolve alongside technological capabilities, with distributed computing and real-time data processing enabling increasingly sophisticated interaction models [8].

These technological advances will continue to reshape the marketing landscape, enabling increasingly sophisticated approaches to customer engagement that balance personalization with privacy considerations. The research on marketing analytics for data-rich environments published in the Journal of Marketing provides a forward-looking perspective on how analytical capabilities will evolve to support enhanced marketing effectiveness. The framework establishes that analytics applications will become increasingly embedded within marketing workflows, transitioning from isolated initiatives to integrated capabilities that inform all customer interactions. The research further explores how analytical sophistication will continue to advance along key dimensions, including data integration, modeling complexity, activation speed, and decision automation. The study specifically addresses how privacy considerations will influence analytical approaches, with increased emphasis on consensual data collection and transparent utilization practices [7].

7. Conclusion

The convergence of big data technologies and distributed computing frameworks within customer data platforms represents a paradigm shift in marketing capabilities. These platforms enable organizations to create unified customer views across channels and activate this data in real-time, transforming marketing from segment-based communications to individually tailored experiences. The research demonstrates that CDP implementations enhance marketing effectiveness through more precise targeting, relevant messaging, and consistent cross-channel experiences. They simultaneously address the growing tension between personalization expectations and privacy concerns by leveraging consensual first-party data rather than third-party tracking. As customer expectations for personalized experiences continue to rise, distributed computing infrastructure becomes increasingly critical for competitive marketing operations. Looking forward, the integration of artificial intelligence, privacy-preserving computation, and edge processing will further enhance marketing capabilities, enabling more predictive, automated, and contextually aware customer engagement.

References

[1] Weiguo Fan, et al, "The Power of Social Media Analytics," June 2014, Communications of the ACM, Online, Available: https://www.researchgate.net/publication/259148570_The_Power_of_Social_Media_Analytics

[2] Salesforce, "State of the Connected Customer," Sixth Edition, Online, Available: https://www.paypalobjects.com/marketing/web23/shared/campaign/enterprise/think-forward-2024/State-of-the-Connected-Customer.pdf

[3] Lauren Hofmann, et al, "The Total Economic Impact™ Of Segment's Customer Data Platform Within The Retail Industry," AUGUST 2021, FORRESTER, Available: https://gopages.segment.com/rs/667-MPQ-382/images/TEI__Segment-Customer-Data-Platform_FINAL_8.9.21.pdf

[4] MATEI ZAHARIA, et al, "Apache Spark: A Unified Engine for Big Data Processing," NOVEMBER 2016, Available: https://people.eecs.berkeley.edu/~matei/papers/2016/cacm_apache_spark.pdf

[5] P.K. Kannan, et al, "Digital marketing: A framework, review and research agenda," International Journal of Research in Marketing, Volume 34, Issue 1, March 2017, Available: https://www.sciencedirect.com/science/article/abs/pii/S0167811616301550

[6] Arne De Keyser, et al, "A Framework for Understanding and Managing the Customer Experience," January 2015, Available: https://www.researchgate.net/publication/336916723_A_Framework_for_Understanding_and_Managing_the_Customer_Experience

[7] Michel Wedel, et al, "Marketing Analytics for Data-Rich Environments," May 2016, Journal of Marketing, Available: https://www.researchgate.net/publication/303482496_Marketing_Analytics_for_Data-Rich_Environments

[8] Colleen Harmeling, et al, "Toward a theory of customer engagement marketing," December 2016, Journal of the Academy of Marketing Science, Available: https://www.researchgate.net/publication/311689280_Toward_a_theory_of_customer_engagement_marketing