



CHECKLIST REPORT

2017

Six Strategies for Advancing Customer Knowledge with Big Data Analytics

How a comprehensive, 360-degree view of customers based on a spectrum of data can enrich actionable insights

By David Stodder

Sponsored by:

cloudera

tdwi
Transforming Data
With Intelligence™

OCTOBER 2017

TDWI CHECKLIST REPORT

Six Strategies for Advancing Customer Knowledge with Big Data Analytics

How a comprehensive, 360-degree view of customers based on a spectrum of data can enrich actionable insights

By David Stodder



555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org

TABLE OF CONTENTS

- 2 **FOREWORD**
- 3 **NUMBER ONE**
Bring data together to gain more complete views of customers
- 4 **NUMBER TWO**
Utilize technologies that enable fast, agile user interaction with big data
- 5 **NUMBER THREE**
Use big data analytics to learn about the customer journey
- 6 **NUMBER FOUR**
Apply customer knowledge to personalize marketing and interaction
- 7 **NUMBER FIVE**
Improve customer experiences across channels using big data analytics
- 8 **NUMBER SIX**
Govern customer data to ensure security and respect for privacy
- 9 **A FINAL WORD**
- 9 **ABOUT OUR SPONSOR**
- 10 **ABOUT THE AUTHOR**
- 10 **ABOUT TDWI RESEARCH**
- 10 **ABOUT TDWI CHECKLIST REPORTS**

© 2017 by TDWI, a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or in part are prohibited except by written permission. Email requests or feedback to info@tdwi.org.

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies. Inclusion of a vendor, product, or service in TDWI research does not constitute an endorsement by TDWI or its management. Sponsorship of a publication should not be construed as an endorsement of the sponsor organization or validation of its claims.

FOREWORD

It is a competitive advantage to know more about your customers and to apply this knowledge to marketing, sales, support, and the development of products and services. Though some organizations previously had experience with statistics and data mining on a limited scale, the advent of big data and the expansion of data science, predictive analytics, and artificial intelligence capabilities (such as machine learning) have transformed the way organizations think about how to gain knowledge about customers and apply it to key initiatives.

No longer limited to just a few sources of transactional data and customer relationship management (CRM) records, organizations today can analyze behavioral data trails left by customers as they move through multiple channels, including social media. In addition, more customers today are using mobile devices to research and shop for products; organizations need to include geolocation data and information generated by customers' use of mobile applications as part of their big data view.

By gathering together the assortment of big data available to them and applying advanced analytics and data science techniques, organizations can gain a detailed, contextual understanding of customers' paths to purchase, what types of marketing strategies are most effective, and how customers influence—and are influenced by—other customers. Increasingly, organizations can implement visualization tools supported by broadly accepted, open source–based data management technologies to enable not just trained analysts and data scientists but business users to interact directly with big data.

Whether aided by data scientists or on their own, business users can apply analytic insights to improve the efficiency and effectiveness of marketing campaigns and enhance customer satisfaction across channels. A more complete big data analytics perspective can lift organizations above siloed perspectives to see a single, 360-degree customer view. This more complete perspective can be the basis for holistic, omnichannel customer marketing and engagement strategies that are superior to disparate strategies based on siloed views.

This TDWI Checklist focuses on six strategies for advancing customer knowledge with big data analytics. It begins with the all-important first step: gaining as close to a complete, 360-degree view of customers as possible. Big data platforms that implement open source Apache Hadoop technologies, which now include systems and code libraries developed using the Apache Spark framework, enable organizations to assemble data for a 360-degree view. The Checklist then explores how to expand the impact of big data analytics while applying governance to ensure proper care of customer data. Taken together, the six strategies will help you apply big data analytics to attracting and retaining your organization's most valuable asset: its customers.



NUMBER ONE

BRING DATA TOGETHER TO GAIN MORE COMPLETE VIEWS OF CUSTOMERS

When decision makers can only see fragments of customer data at a time, it limits their understanding of both individual customers and larger trends across populations of customers and prospects. Organizations dependent on such fragmentary views often cannot be agile with information because the limited sources support only one or a few types of queries and reports. Data science and advanced analytics deliver more value if organizations can deploy models and algorithms on detailed data brought together from many sources, not just on summaries or samples contained in a single data warehouse or data mart.

Bringing together as much of this data and information as possible will help organizations gain a richer, more detailed picture of customers. As their intelligence grows from the newly combined data, organizations can add insights to customer profiles, increasing their value exponentially. These enhanced profiles can improve real-time customer engagement, sharpen marketing campaigns, and support use of analytics to predict customers' future behavior. For these reasons, it is important for modern organizations to seek a comprehensive, 360-degree view of customers and prospects.

The best way to achieve this is to bring customer data together into a single, central data repository or enterprise data hub that gathers together diverse data generated by customers in different channels and at different touchpoints. This is not trivial to do in a data warehouse because customer data is often held in disconnected data silos or applications that have different data types and models. Customer transaction records are in one or several databases linked to OLTP systems or business applications, while CRM and sales force management systems hold other bits of vital data. Then there are the even less structured customer service records, call center interaction records, customer satisfaction survey data, and information generated by online behavior and engagement. External social media data could also be valuable for analysis.

Although collecting all this information in a classic data warehouse using a single structured data model might seem ideal, in implementation data warehouses are difficult and slow to create. They are usually not comprehensive enough and have limitations in terms of the types of analytics that they can support. Some organizations try to use data virtualization or federation middleware to expand the data reach by creating virtual views across sources. This option could work for accessing some additional data sources, but virtualization

can require complex up-front modeling and setup and does not enable organizations to access the full mix of customer information at the file level.

With the maturity of the Hadoop and Spark ecosystem, in which organizations can now include the Apache Impala massively parallel processing (MPP) database engine, many organizations are choosing to collect big data in an open, less model-dependent enterprise data hub. The hub can hold large volumes of diverse data and use commodity clusters and other ecosystem technologies to be horizontally scalable to meet business demand.

Organizations need a strategy for creating the single customer view. Even if it is not possible to move absolutely all of the data and information into a single source such as an enterprise data hub, the more data they can locate in this source, the sooner they will be able to decrease time to insight through application of big data analytics.

¹ For a more detailed definition of an enterprise data hub, see <http://searchcio.techtarget.com/definition/enterprise-data-hub>

 **NUMBER TWO**
UTILIZE TECHNOLOGIES THAT ENABLE FAST, AGILE USER INTERACTION WITH BIG DATA

Agility is a critical success factor for customer data management. This is because use cases for customer data can range widely, from simple consumption of reports and metrics for operational users in sales and service, to visual data discovery and advanced data science for personalized marketing and development of new products and services. Organizations need to evaluate data management technologies that can support diverse use cases. In addition, as more business users move beyond just consuming data, data management platforms must allow for more data interactivity, faster iterations to test models and examine detailed data, and access to fresher data.

Here, we highlight five technologies that organizations should consider when building a strategy for higher speed and agility with big data customer analytics:

- **Columnar databases.** A columnar database management system stores data in columns instead of rows, which for many types of analytics can offer faster and more efficient query performance. These databases are most effective when reading a limited number of columns because queries do not have to access entire rows as they would in a traditional row-based database. In the Apache Hadoop and Spark ecosystem, two open source–based columnar systems to evaluate would be Hbase and Kudu.
- **In-memory databases.** Data platforms today can take advantage of large server memory, enabling users to search and access more data without having to retrieve it from disk. In-memory databases can reduce the amount of data preparation necessary and provide decent performance in instances such as building indexes. As a result, users are encouraged to experiment and be agile rather than worry about the possible performance cost. In-memory database technology can provide users with faster interactive access to data for analytics, including testing of models.
- **In-database computing.** Moving data from one platform to another, particularly for data preparation and transformation, can slow overall performance. With the volume of data required for analytics, being able to run more processes where the data resides in the hub can improve speed.
- **Machine learning for data science and preparation.** Machine learning is critical for faster discovery of patterns and relationships in large volumes of data and for building predictive insights. Along with its use for analytics, machine learning can help automate traditionally lengthy data preparation steps

such as profiling, cleansing, and cataloging. Organizations should evaluate solutions that use machine learning to reduce the time required by users and data scientists to prepare data.

- **Cloud computing.** The cloud is becoming a mainstream option for many computing and storage requirements. In addition, some data that is important for customer analytics is native to the cloud because it is generated by cloud-based services. Organizations can take advantage of cloud computing to quickly spin up a data system to support analytics for meeting an immediate business need such as a marketing campaign. Rather than extract and load data from the cloud into an on-premises data warehouse or wait for on-premises systems to be available, organizations can set up cloud-based sandboxes to explore data, run predictive algorithms, or perform “what if” simulations and scenarios.



NUMBER THREE

USE BIG DATA ANALYTICS TO LEARN ABOUT THE CUSTOMER JOURNEY

Customers begin a journey from the moment they first interact with your organization as potential prospects. The interactions could occur across one or more channels, including different online channels. The journey might begin with a click on a search engine result, a visit to a website, entrance into a physical store, or through contact with a call center agent. Knowing what behavior prompted this initial interaction could not be more valuable, especially when known not just about a single customer but also about larger segments of similar customers. This knowledge can help organizations build better, more comprehensive customer records.

From the initial point of contact, data generated during a customer's journey is a rich source of knowledge that can improve marketing, service, product development, and overall strategy. The drive to understand the customer journey (also called a customer life cycle) can provide context for collecting the diverse data that goes into creating a 360-degree view of the customer. However, data generated by interactions across multiple channels is typically stored in disparate and disconnected data silos. If this data is gathered into a central data hub, organizations can begin to fill in gaps in their knowledge about the customer journey.

As they gather data, organizations can analyze it for trends, patterns, and correlations that help them understand attributes of buying behavior and what influence marketing has on customers' paths toward a purchase decision. Organizations can then run predictive models to derive insights that can help marketing and service functions to achieve better outcomes and higher customer satisfaction. These insights are highly time-sensitive, though. By collecting the data into a central hub, organizations can cut down on the often time-consuming steps required to query data and run analytic models across multiple, disparate silos (assuming that you have access to the data in those siloed systems in the first place).

Through understanding the customer journey, organizations can address more intelligently the problem of churn—that is, the moment when customers end their journey, often to go to a competitor. By analyzing the rate of churn and predicting when it might occur, organizations can determine how and when to act to dissuade customers from leaving. They can also use analytics to identify their most important and profitable customers so that they can focus on them and anticipate behavior that is indicative of churn. Organizations can then be proactive in taking actions to keep customers engaged.

Organizations should focus on collecting data generated at all points during a customer's journey into a hub that can serve as the source for a variety of analytics for understanding and predicting retention, attrition, and the rate of churn. Ultimately, the analytics will enable organizations to understand the entire journey and optimize marketing and service actions to fit each stage of the life cycle.



NUMBER FOUR

APPLY CUSTOMER KNOWLEDGE TO PERSONALIZE MARKETING AND INTERACTION

Big data analytics is revolutionizing how organizations personalize marketing offers and shape customers' experiences. The volume of behavioral data generated by customer activity across channels can give marketers a tremendous resource for gaining a detailed view of customers' historical and real-time behavior, which organizations can analyze to generate predictive insights into how customers will respond to future offers.

The possibility of smarter personalization is a major reason for collecting data into an enterprise data hub. With a 360-degree view of customer behavior, organizations can apply analytics to reveal unexpected data relationships and associations. This knowledge can give organizations not only more accurate views of their customers, but also insights that competitors do not have to fuel differentiating personalization strategies.

Online retailers rely on personalization to drive recommendation engines, which can automatically deliver offers based on data associations identified through analysis of customer data, such as purchase history and clickstreams generated during website visits. Marketing functions can use big data analytics to expand their contextual understanding of the impact of personalized offers and to optimize when to make the offers to achieve the best outcomes.

Data-generating events to analyze include responses to email offers, website behavior, the customers' location (particularly if the customer is using a mobile device), and social media activity. Using big data analytics with machine learning, organizations can develop deeper and more detailed knowledge of customers to enable them to deliver targeted offers appropriate for different stages of the customer journey and for each different type of channel.

Here are four recommendations for using big data effectively to improve personalization:

- **Ensure that big data analytics is tightly integrated with e-commerce and marketing campaign engines.** Too often, latency exists between analytics and action. Personalization of offers is most effective if the offers are provided to customers in near or true real time.
- **Build contextual understanding for comprehensive insights into customer behavior.** Though not every correlation or other association found in the data is relevant, organizations can benefit from using analytics on varied, usually unstructured, big data to build contextual understanding around the structured transactional data captured by OLTP systems.

- **Test analytics carefully before using results to drive personalization campaigns.** No organization wants to irritate customers with irrelevant offers or scare them by “knowing too much”—that is, exhibiting a lack of respect for privacy. Test personalization on smaller groups to ensure that programs are appropriate. Apply data science to understand the reverse causal effects of targeting.
- **Evaluate whether your data platform can handle the performance requirements.** Personalization operationalizes analytics. If marketing and recommendation engines have to wait for data platforms to produce results, systems will be too inefficient and organizations will miss opportunities. Ensure that your data platform can handle performance and availability requirements.



NUMBER FIVE

IMPROVE CUSTOMER EXPERIENCES ACROSS CHANNELS USING BIG DATA ANALYTICS

For most organizations, it is far more cost-effective to keep current customers loyal than to replace them with new customers. A critical factor in increasing customer loyalty and growing market share is providing the best customer experience. Organizations can apply big data analytics to learn what customers want and be proactive in tuning customer experiences to ensure loyalty.

In particular, big data analytics can help organizations interpret customer behavior faster so they can prepare human agents or automated systems to respond in ways that increase satisfaction. Through dashboards and other types of visualizations, analytics can be made actionable so that sales, service, and support personnel are informed as they engage with customers.

To improve online customer experiences, big data analytics can help organizations identify and apply personalization insights not only to automate recommendation systems but also to improve the content delivery and visual design of websites to optimize and personalize customer experiences.

When organizations make it a priority to use analytics to improve customer experiences, it puts pressure on the personnel responsible for big data analytics to deliver insights that are timely, relevant, and appropriate for the context. They need technologies that will help them collect, prepare, and analyze data in time to make a difference. This is a major reason why many firms are investing in data science and the development of machine learning algorithms that can identify patterns in both historical and real-time data rapidly to produce insights that can be applied at the point of decision. Organizations can use predictive analytics to anticipate what customers will do if their experience is tailored with certain types of content, offers, or different styles of communication.

One of the biggest challenges in understanding and affecting customer experience is the multichannel nature of customer interactions. Commerce occurs mostly in discrete events online, in physical stores, on mobile devices, and through contact center interactions. Each of these touchpoints has its own applications, styles of engagement, and metrics for measuring quality. It can be difficult to gain a big-picture view of a single customer's experience, much less a population or segment of customers.

Using the concept of the customer journey (or life cycle) can be helpful in creating a context for improving customer experiences. Organizations can then apply path analysis to identify and interpret the series of experiences that might precede a decision to buy or not buy a product or service. Organizations can analyze where a lack of quality in the experience may have caused a customer to churn. Path

analysis can use data collected in an enterprise data hub to look for patterns and correlations in data generated from multiple channels. Analytics can be used to examine detailed data about customer segments or affinities. Marketing decision makers can use analytics to develop an omnichannel marketing and engagement strategy that sets out how the organization will coordinate actions across all channels, not just each one separately.

Organizations should ensure leadership support for identifying which data sources are most valuable and how to use analytics to improve customer experiences. This leadership is essential for gaining buy-in from managers in sales, service, and support functions. Through such leadership, analytics professionals can collaborate with business managers to address interpersonal and political challenges that sometimes arise when attempting to apply analytics insights to change sales, service, and support personnel behavior. Communication is also key to gaining feedback about what worked and what did not in applying the analytics in real-world sales, service, and support.

 **NUMBER SIX**

GOVERN CUSTOMER DATA TO ENSURE SECURITY AND RESPECT FOR PRIVACY

Governance should be part of any modern data and information management strategy, but it is especially important for organizations that are analyzing customer or patient data. Data governance is about creating plans, policies, rules, and accountability for overseeing data—both data at rest and as it moves through and beyond the organization. Thus, governance must cover how the organization protects customer data, including personally identifiable information (PII) that falls under industry and government regulations.

Personnel responsible for data governance should understand how to follow directives such as the European Union’s emerging General Data Protection Regulation (GDPR) or for U.S. healthcare organizations analyzing patient data, the Health Insurance Portability and Accountability Act (HIPAA).

Organizations want to avoid financial penalties, as well as damage to their reputations, that can result from poor data protection and inadequate privacy practices. Customers can lose trust if their privacy is violated, either through data breaches or poorly thought-out use of their data for marketing. Thus, it is important to ensure effective governance at every phase in the customer journey from the ingestion of data to its application.

Because customer data originates from multiple sources, governance of big data analytics can be complicated. Organizations need to find the right balance between allowing data science and exploration essential to innovative analytics and necessary adherence to privacy policies, regulations, and customer expectations. Here are four recommendations:

- **Consider customers’ concerns about privacy in marketing strategies.** Organizations must ensure that customers will not react negatively to perceived breaches of their privacy in marketing campaigns. Governance should document best practices for marketing, sales, service, and social media functions to follow before applying big data analytics so that organizations can change potentially unethical or discomfoting strategies before launching a marketing campaign.

- **Facilitate business/IT collaboration on governance.** Although in most organizations IT is primarily responsible for governance, it is more effective if business executives are part of designated committees and champion its importance. Such committees can provide joint business/IT leadership on related matters of data stewardship, training, and project prioritization.
- **Adjust policies and practices to fit development of big data hubs.** As organizations increase their use of data hubs to collect customer data from diverse sources, they need to adjust governance practices that may have been set up to work with traditional data warehouses that only manage strictly sourced and structured data. Organizations should examine security and governance practices with data hubs to ensure that they are using available technologies effectively to standardize and automate data management. Organizations should also ensure that personnel follow rules for data masking and anonymization to protect customers’ or patients’ identities.
- **Support governance with master and metadata management.** Data governance is strengthened by good knowledge of and documentation about the organization’s data and how it is being used. Master data and metadata management, which are helpful in establishing higher-level definitions of customers and organizing data related to them, can also help organizations govern big data.

A FINAL WORD

With the right technologies and practices, big data analytics can help organizations make rapid strides in gathering comprehensive knowledge about customers and using predictive insights to improve satisfaction, loyalty, and ultimately, profitability. Big data analytics can give organizations unique competitive differentiation, particularly in how they personalize marketing campaigns and provide cross-sell and upsell recommendations. Your competitors will not know your customers as well as you do. The six strategies described here are important to gaining a 360-degree view of customers and extending the benefits of well-governed big data analytics to the application of customer knowledge for a diverse range of marketing and engagement initiatives.

ABOUT OUR SPONSOR

cloudera

www.cloudera.com

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing, and analyzing vast amounts of data. Learn more at cloudera.com.

ABOUT THE AUTHOR



David Stodder is senior director of TDWI Research for business intelligence. He focuses on providing research-based insights and best practices for organizations implementing BI, analytics, data discovery, data visualization, performance management, and related technologies and methods and has been a thought leader in the field for over two decades. Previously, he headed up his own independent firm and served as vice president and research director with Ventana Research. He was the founding chief editor of Intelligent Enterprise where he also served as editorial director for nine years. You can reach him by email (dstodder@tdwi.org), on Twitter ([@dbstodder](https://twitter.com/dbstodder)), and on LinkedIn (linkedin.com/in/davidstodder)

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data management solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

ABOUT TDWI CHECKLIST REPORTS

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.