

Matt Ariker, Nimal Manuel,  
and Marc Singer

## The big data trinity: Creating an analytics system to support a learning culture

Marketing & Sales July 2015

As companies look to build up their “test and learn” capabilities, they need to create a 3D-360° understanding of the customer, a plan for building insight models, and a self-learning ecosystem that scales.

**“Test and learn” has emerged as a tried and true method** for an organization to get closer to its customers. This approach is helping companies move more quickly, incorporate real feedback from customers, and be agile enough to adapt.

While “test and learn” has plenty of examples of success, most of them are isolated to a few small “swat” teams empowered to work separately from the organization, so the benefits are limited. To scale a test-and-learn capability requires a more fundamental re-architecting of a company’s big data and analytics capabilities.

In our experience, this requires putting in place a “big data trinity”:

- A 3D-360° understanding of the customer
- An analytics roadmap designed to define the analytic strategies and their embedded respective data requirements

- A complete and integrated analytics ecosystem that connects insights to actions and results as part of a constant, self-improving cycle

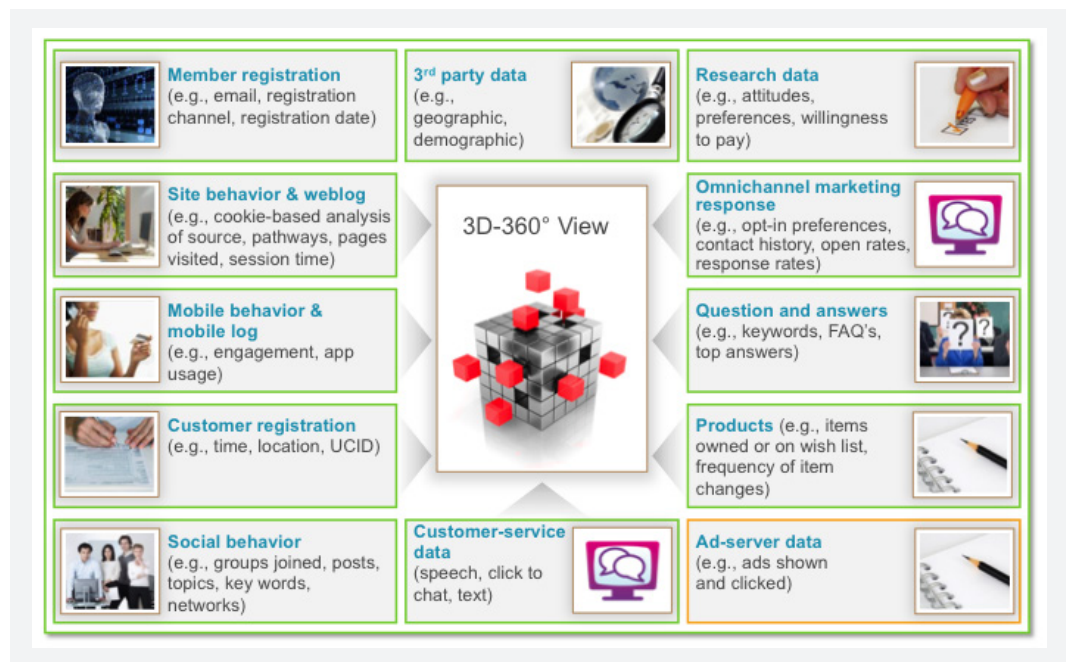
### 1. A complete 3D view of the customer

Any learning system at scale needs to start with a data-driven approach to understanding the customer. Given the complexity of coaxing meaning from this data, companies tend to focus on the structured data—common sources of data that fit in prescribed models, such as billing and order-processing information—that they can easily access to build up a 360° view of their customer. Adding to the difficulty is that “unstructured” data (e.g. text, web, social media, location-based data, images, video, sensor, and speech) are both notoriously difficult to analyze and often live in different databases from structured data, leading to uncoordinated analyses. Unfortunately this leaves businesses with an incomplete “2D” customer view that is often also incorrect.

But companies are missing the bigger picture, literally. The more insightful (and valuable) 3D-360° view is based on a combination of a customer’s structured data (who, what, when) and unstructured data (why and nuance) (exhibit 1).

#### Exhibit 1

3D-360° customer view includes structured and unstructured data



The objective (and the tricky part, of course) in combining structured and unstructured data into a 3D-360° customer view is to get a clearer picture of intention and context. For example, a bank might have a deep view of a mass-market, affluent customer based

on previous transactions. If this customer were to tweet about the birth of twins and post new pictures online, the bank could then consider offering life insurance through its other holdings.

To make this merger work requires companies to do two things:

- Create a business-specific semantic architecture to provide context and meaning. For example, words (spoken, written, or typed) need at least (1) a dictionary to understand the individual meaning of the word, including synonyms and antonyms, and (2) a taxonomy to understand meaning in the context of words used together. This semantic architecture helps provide the context so very important to understanding, making it a clear best practice for analyzing unstructured data.
- Link unstructured data to structured data. Once you have context and meaning for your unstructured data, they need to be linked to the source and existing structured data. For example, in the case of a customer comment, the process might look something like this: “This comment was sent from these GPS coordinates with this frequency and duration. This comment is associated with this device or machine and this person, who is a customer, and whose annual spend with us is X amount.” Automated processes can help with the linkages, but it still requires managerial oversight and good quality-assurance processes to confirm accuracy.

## 2. A plan for building insights

An analytic roadmap is a plan for defining what analytics capabilities need to be created and when in order to support an ongoing, full-scale learning program (exhibit 2).

### Exhibit 2

Analytics roadmaps have six dimensions



Creating a good roadmap requires three steps. The first one is identifying what questions you want to answer. For example, if the strategy is to drive margins by acquiring more high-value lifetime customers, a question might be “What sorts of tweets and blogs do people in this segment post?”

Determining how to answer those questions leads to the second step: building out analytics requirements. For the sample question above, this would include systems that combine social media with regular data, analytics tools to interpret text, relevant sources of data, software that ties social media actions to personal IDs in a CRM system, customer-decision-journey capabilities, and data-integration services, to name a few.

Being specific about the question is crucial, because even seemingly unimportant variations can have drastic implications. For example, understanding what high-value customers blogged about or tweeted 15 minutes ago requires rapid metrics-reporting capabilities and fluid data dynamics, which in turn require agile analytic solutions that can link in real time to implementation engines. With a system like this in place, you can then ask many more questions that require those same capabilities. For example, this system would also be helpful in identifying low-value customers by their tweets and blogs, enabling companies to avoid wasting resources on acquiring them.

The third step is to sequence the build-out of the analytics capabilities and their associated use cases—typically eight to ten over an 18-month period—that can answer the identified questions. This sequencing defines the roadmap and has important implications when it comes to operating-expense and capital-expense requirements. Different combinations of analytics software have different requirements for data, storage, and cloud, which in turn drive different total cost of ownership from build (CapEx) and run (OpEx) perspectives.

In sequencing these efforts, companies should focus on building out the less complex analytics capabilities first. Doing so requires analysis to develop a clear understanding of trade-offs in terms of cost, time, and completion risk. Given the realities of a demanding organization, for example, the most important requirement may be the ability to deliver results quickly to build a convincing case for additional resources to develop additional models.

One important thing to note about analytics roadmaps is that their impact is as dependent on people as on data and tools. Companies need to find not only the people who can manage the software and tools, but data scientists to do the analytics and “business translators” who can turn insights into business value.

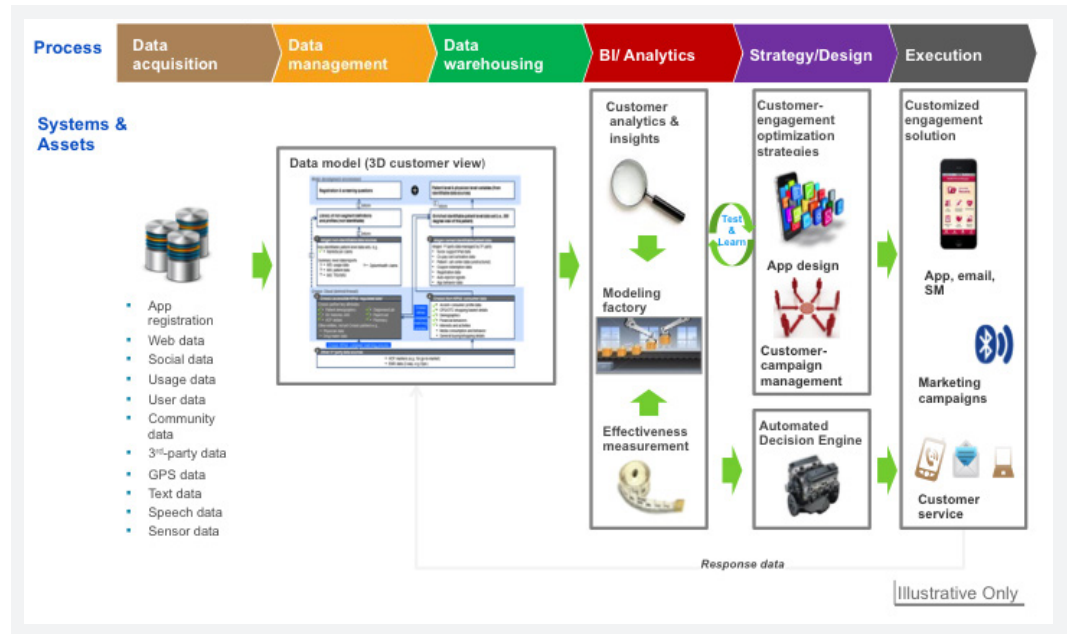
### **3. Self-learning ecosystem**

Building a capability that can learn and react quickly requires the integration of data discovery and analysis, insights delivery, and campaign execution. This in turn requires hand-off

processes as well as middleware and APIs that tie systems together to allow data and decisions to move quickly through them (exhibit 3).

### Exhibit 3

End2end analytics ecosystem



But what's critical for organizations that want to drive growth is that a test-and-learn capability go far beyond standard data models that are optimized to simply deliver messages or offers; it's actually a two-way street that includes collecting and tracking the customer's reactions to the offer so the company can learn and improve future interactions. Technology has advanced to such a degree that much of this cycle can be automated and architected to learn and adapt, for example, to personalize a landing page in real time. Driven by automation, self-learning has the ability to quickly scale, allowing a company to create ten times the number of insights that it does today.

Self-learning requires a different design approach from more typical systems, which focus on pushing things through and then out of the system. For one thing, the system needs technologies to track how people respond to messages and offers across a vast array of channels (email, mobile web, app, in-store, website, social media, etc.) and then a series of virtual "pipes" to feed that information back into the 3D view of the customer to develop better statistical and event-based models (e.g., response rates based on an event and context). That mechanism needs to be constantly updated as well, since the number of digital touchpoints is expanding by about 20 percent every year.<sup>1</sup>

<sup>1</sup> "Brand success in an era of Digital Darwinism," *McKinsey Quarterly*, 2015

The problem is that while companies can use machine learning to quickly scale insights, it's worthless if the organization cannot act on each of the new insights. The sheer mass of insights now possible makes this an all-too-real issue for most companies. Addressing the problem requires putting in place a range of supporting capabilities to make the program work. One example might be rules to help automate decisions, such as what is the lowest cost for an offer, which messages to send to a given person who hasn't seen an offer before, etc. More complex decision-making rules, exceptions, or unacceptable variances could be programmed to escalate to managers to handle, though such exceptions should be less than 5 percent of all decisions. These rules and guidelines should be stored in a library that systems can access as needed.

In the same way, companies need to develop content libraries to provide a ready supply of text and images to send to customers. If users are going to see a personalized message specifically designed for them on a website, for example, someone needs to design the text and images.

Authors of a recent study in *Nature* highlight how this approach is applied to robotics via an "intelligent trial-and-error algorithm" that guides how robots learn. If a robot realizes it isn't moving the way it ought to, for example, it tests other ways of getting where it needs to go. But its choices are based on an extensive database of 30,000 possible movements to which it has access—in other words, a library of options.<sup>2</sup>



Discovering insights on an individual basis and then being able to act on them quickly is becoming a cornerstone of advanced analytics leveraging big data. Pulling together all the elements of this kind of system takes commitment and discipline. But it may be the only option for companies that want to accelerate their growth. □

<sup>2</sup>"Unbreakable: A Robot That Can't Be Stopped," *The Atlantic*, May 2015

**Matt Ariker** is the Chief Operating Officer of McKinsey's Consumer Marketing Analytics Center (CMAC); **Nimal Manuel** is a principal in McKinsey's Kuala Lumpur office, and **Marc Singer** is a director in the San Francisco office.