Chapter 3
Big Data Analytics Applications

This chapter discusses a number of important use cases for Big Data Analytics. In each case, Big Data Analytics is becoming integrated with business processes and traditional analytics to provide major outcomes. In many cases, these use cases represent game changers essential to the survival and growth of an organization in an increasingly competitive marketplace. Some of these use cases are still in their infancy, while others are becoming increasingly commonplace.

3.1 Social Media Command Center

Last year, Blackberry faced a serious outage when its email servers were down for more than a day. I tried powering my Blackberry off and on because I wasn't sure whether it was my device or the CSP. It never occurred to me that the outage could be at the Blackberry server itself. When I called the CSP, they were not aware of the problem. For a while, I was okay without receiving any emails, but then I started to become curious. So I turned to one obvious source: Twitter. Sure enough, I found information about the Blackberry outage on Twitter.

One of my clients told me that his VP of Customer Service is glued to Twitter looking for customer service problems. Often, someone discovers the problem on Twitter before the internal monitoring organization. We found that a large number of junior staffers employed by marketing, customer service, and public relations search through social media for relevant information. Does this sound like an automation opportunity?

A *Social Media Command Center* combines automated search and display of consumer feedback expressed publicly on the social media. Often, the feedback is summarized in the form of "positive" or "negative" sentiment. Once the feedback is obtained, the marketer can respond to specific comments by entering into a

Figure 3.1: Gatorade Social Media Command Center

conversation with the affected consumers, whether to respond to questions about an outage or obtain feedback about a new product offering.

The marketing organization for Gatorade, a sports drink product, decided to create a Social Media Command Center to increase consumer dialog with Gatorade.[12] Figure 3.1 shows the monitoring station with the dashboard. Big Data Analytics can be used to monitor social media for feedback on product, price, and promotions as well as to automate the actions taken in response to the feedback. This may require communication with a number of internal organizations, tracking a product or service problem, and dialog with customers as the feedback results in product or service changes. When consumers provide feedback, the dialog can only be created if the responses are provided in low latency. The automated solutions are far better at systematically finding the information, categorizing it based on available attributes, organizing it into a dashboard, and orchestrating a response at conversation speed.

3.2 Product Knowledge Hub

As consumers turn into sophisticated users of technology and the marketplace becomes specialized, the product knowledge seldom belongs to one organization. Take the Apple iPhone as an example. The iPhone is marketed by Apple, but its parts came from a large supply chain pool, the apps running on the iPhone come

from a large community of app developers, and the communications service is provided by a CSP. Google's Android is even more diverse, as Google provides the operating system while a cell phone manufacturer makes the device. The smartphones do not work in isolation. They act as WiFi hubs for other devices. So, what happens if I want to know how to tether an iPhone to an Apple iPad? Do I call my CSP, or do I call Apple? Would either of their websites give me a simple step-by-step process I can follow?

Every time I get into these technical questions about products I am trying to use, I end up calling my son, who happens to know the answers to any such question. Recently, he decided to educate me on how he finds the answer, and so I was introduced to a myriad of third-party sites where a variety of solutions can be found. In most cases, we can find them by searching using any popular search engine. However, the solutions do not always favor the CSPs, and they are often dated, failing to take into account the latest offerings. Between the device operating system, the offerings from CSPs, and the apps, one must tread carefully through the versions to make sure the solution we discover is for the same version of software that is on the device. So now, we are facing data that is characterized by both variety and veracity. Can we use Big Data Analytics to solve this problem?

The solution involves three sets of technologies. Fortunately, Vivisimo has packaged these technologies into its Velocity product, making it easier to obtain an integrated solution. The first part of the solution is the capability to tap any sources of data. A CSP may already have pieces of the solution on its intranet, put together by product managers or customer service subject matter experts. Or, the information may reside on a device manufacturer site or a third-party site. All this data must be pulled and stripped of its control information so that the raw text is available to be reused.

The second part of the solution is to create a set of indices so that the raw information can be categorized and found when needed. Because many combinations of products exist, we would like to collect and combine information for the devices searched. The federated indexing system lets us organize the information for easy access.

The third part of the solution involves creating an XML document against a query that can either be rendered by a mashup engine or made available to a third-party application.

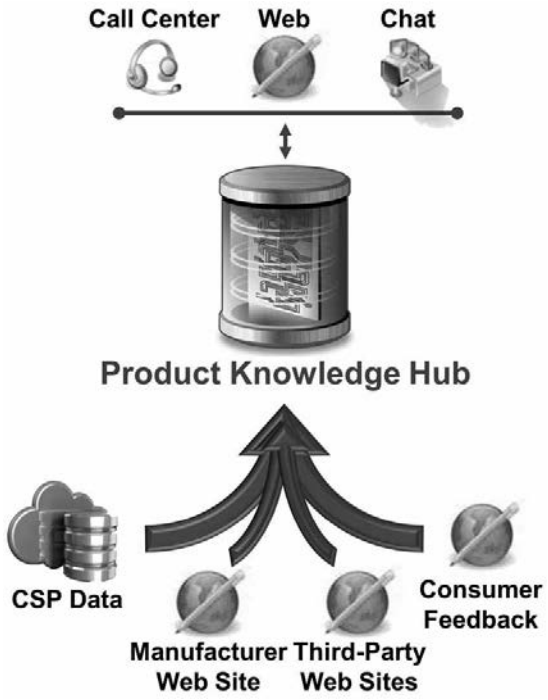

Figure 3.2: Product Knowledge Hub for a CSP

What we have created is a *knowledge hub*, which can now be used directly from a website or made available to the call centers. It significantly reduces call-handling time in the call centers and also increases first call resolution. By placing the information on the web, we are now promoting the CSP's website as the source of knowledge, which increases web traffic and reduces the number of people who resort to contacting the call center. Figure 3.2 depicts the Product Knowledge Hub.

Once we have created a single source of knowledge, this source can be used to upsell other products, connecting usage knowledge to product features and using the knowledge pool to discover new product or business partnership ideas. A lot of stray, fragmented knowledge about the products may be rapidly organized and find a variety of other uses.

3.3 Infrastructure and Operations Studies

A number of industries are exploring the use of Big Data to improve their infrastructure. In many situations, the best way to improve the infrastructure is to

understand its use and how bottlenecks or configurations impact performance. In the past, this data required extensive manual data collection costs. Big Data provides a natural source of data with minimal data collection costs. I will lay out examples from public services to illustrate this point.

The city of Boston decided to use Big Data to identify potholes in the streets by sponsoring a competition in the analyst community. A winner came from Sprout & Co., a nonprofit group in Somerville, Massachusetts. The solution included the use of magnitude-of-acceleration spikes along a cell phone's z-axis to spot impacts, plus additional filters to distinguish potholes from other irregularities on the road. The new algorithm made Street Bump, a free download in Apple's App Store, a winner.[13] This analysis can save significant road survey cost. Navigation systems can also use the cell phone data to avoid traffic congestion and offer alternate routes. This type of use of Big Data is one of the best ways to gain acceptance without getting into privacy or security issues.

In another example, city bus and train agencies are making their real-time transit information available to riders. This information significantly improves the user experience and reduces the uncertainty associated with both planned and unexpected delays. Transloc (*www.transloc.com*) provides this information for riders using a variety of technologies, including smartphones, web, and SMS messages. It also provides prediction capabilities on expected arrival time. Once the app is loaded on a smartphone, the rider can use it to accurately estimate travel time and also review the travel route.

IBM's Smarter Cities® initiative is using Big Data in a number of applications directed at city infrastructure and operations. Location data from cell phones can be used to provide raw material for detecting traffic patterns. These patterns can then be used to decide on new transportation projects, to change controls, or to redirect traffic in case of an emergency.

Another important application for Big Data Analytics is public safety. The New York Police Department is using Big Data for crime prevention.[14]

3.4 Product Selection, Design, and Engineering

Product automation provides an enormous opportunity to measure customer experience. We take photos digitally and then post them on Facebook, providing an opportunity for face recognition without requiring laborious cycles in digitization. We listen to songs on Pandora, creating an opportunity to measure what we like or dislike or how often we skip a song after listening to the part of it that we like the most. We read books electronically online or on our favorite

handheld devices, giving publishers an opportunity to understand what we read, how many times we read it, and which parts we look at. We watch television using a two-way set-top box that can record each channel click and correlate it to analyze whether the channel was switched right before, during, or after a commercial break. Even mechanical products such as automobiles are increasing electronic interactions. We make all of our ordering transactions electronically, giving third parties opportunities to analyze our spending habits by month, by season, by ZIP+4, and by tens of thousands of micro-segments. Usage data can be synthesized to study the quality of customer experience and can be mined for component defects, successes, or extensions. Marketing analysts can identify micro-segmentations using this data. For example, in a wireless company, we isolated problems in the use of cell phones to defective device antenna by analyzing call quality and comparing it across devices.

Products can be test marketed and changed based on feedback. They can also be customized and personalized for every consumer or micro-segment based on their needs. Analytics plays a major role in customizing, personalizing, and changing products based on customer feedback. Product engineering combines a set of independent components into a product in response to a customer need. Component quality impacts overall product performance. Can we use analytics to isolate poorly performing components and replace them with good ones? In addition, can we simplify the overall product by removing components that are rarely used and offer no real value to the customer? A lot of product engineering analytics using customer experience data can lead to building simplified products that best meet customer requirements.

To conduct this analysis and predictive modeling, we need a good understanding of the components used and how they participate in the customer experience. Once a good amount of data is collected, the model can be used to isolate badly performing components by isolating the observations from customer experience and tracing them to the poorly performing component. Complex products, such as automobiles, telecommunications networks, and engineering goods, benefit from this type of analytics around product engineering.

The first level of analysis is in identifying a product portfolio mix and its success with the customers. For example, if a marketer has a large number of products, these products can be aligned to customer segments and their usage. We may find a number of products that were purchased and hardly used, leading to their discontinuation in six months, while other products were heavily used and sparingly discontinued.

Once we have identified less-used products, the next analysis question is whether we can isolate the cause of customer disinterest. By analyzing usage patterns, we can differentiate between successful products and unsuccessful ones. Were the unsuccessful ones never launched? Did many users get stuck with the initial security screen? Maybe the identification process was too cumbersome. How many users could use the product to perform basic functions offered by the product? What were the highest frequency functions?

The next level of analysis is to understand component failures. How many times did the product fail to perform? Where were the failures most likely? What led to the failure? What did the user do after the failure? Can we isolate the component, replace it, and repair the product online?

These analysis capabilities can now be combined with product changes to create a sophisticated test-marketing framework. We can make changes to the product, try the modified product on a test market, observe the impact, and, after repeated adjustments, offer the altered product to the marketplace.

Let us illustrate how Big Data is shaping improved product engineering and operations at the communications service providers. Major CSPs collect enormous amounts of data about the network, including network transport information coming from the routers and the switches, as well as usage information, popularly known as call detail records (CDRs), which are recorded each time we use telephones to connect with one another. As the CSP networks grew in sophistication, the CDRs were extended to data and video signals using IPDRs. Most CSPs refer to this usage information as xDRs (where *x* is now a variable that can be substituted for "any" usage information). For larger CSPs, the usage statistics not only are high volume (in billions of transactions a day) but also require low-latency analytics for a number of applications. For example, detecting a fraudulent transaction or abusive network user in the middle of a video download or call may be more valuable than finding out this information the next day. In addition, it is always a strategic driver for CSPs to lay out all the network and usage information on their network topology and geography and use a variety of automated analytics and manual visualization techniques to connect the dots between network trouble or inefficiencies and usage. The analytics provides CSP with a valuable capability to improve the quality of the communication. If every user call is dropping in a particular area that is a popular location for premier customers, it could lead to churn of those customers to competitors.

The information about xDRs, network events, customer trouble tickets, blogs, and tweets in the social media can be correlated for a variety of business

purposes. CSPs have used this analytics to detect spots with poor network performance to reorganize towers and boosters. The differences in usage can be analyzed to detect device problems such as faulty antennas on specific models. The variations can also be analyzed to find and fix network policies or routing problems. As CSPs race to implement high-volume, low-latency xDR hubs, they are finding plenty of business incentives to fund these programs and reap benefits in the form of improved product offerings to their customers.

3.5 Location-Based Services

A variety of industries have location information about their customers. Cell phone operators know customer location through the location of the phones. Credit-card companies know the location of transactions, and auto manufacturers the location of cars, while social media is trying its best to get customers to disclose their location to their friends and family. On a recent short trip to India, I decided to use Endomondo, an app on my cell phone to record my jogging activity in Mumbai, India, which was instantly posted on my Facebook page, thereby letting my friends know of my visit to Mumbai.

Let us take a wireless CSP example to study how we collect and summarize location information. A cell phone is served by a collection of cell phone towers, and its specific location can be inferred by triangulating its distance from the nearest cell towers. In addition, most smartphones can provide GPS location information that is more accurate (up to about 1 meter). The location data includes longitude and latitude and, if properly stored, could take about 26 bytes of information. If we are dealing with 50 million subscribers and would like to store 24 hours of location information at the frequency of once a minute, the data stored is about 2 terabytes of information per day. This is the amount of information stored in the location servers at a typical CSP.

Customer locations can be summarized into "hang outs" at different levels of granularity. The location information can be aggregated into geohashes that draw geo boundaries and transform latitude-longitude data into geohash so that it can be counted and statistically analyzed. The presence of a person in a specific location for a certain duration is considered a space-time box and can be used to encode the hang out of an individual in a specific business or residential location for a specific time period.

Many of our smartphone apps collect location data, provided a subscriber "opts-in."[15] If a marketer is interested in increasing the traffic to a grocery store that is located in a specific geohash, they can run an effective market-

ing campaign by analyzing and understanding which neighborhood people are more likely to hang out or shop in that specific grocery store. Instead of blasting a promotion to all neighborhoods, the communication can now be directed to specific neighborhoods, thereby increasing the efficiency of the marketing campaign. This analysis can possibly be conducted using 6-byte location geohash over a span of one hour and finding all the cell phones that have visited the grocery store regularly. A predictive model can compute the probability of a customer visiting the grocery store based on their past hang out history, and customer residence information can be clustered to identify neighborhoods most likely to visit the shopping center.

Analysis of machine-to-machine transaction data using Big Data technologies is revolutionizing how location-based services can be personalized and offered at low latency. Consider the example of Shopkick, a retail campaign tool that can be downloaded on a smartphone. Shopkick seeks and uses location data to offer campaigns. Once the app is downloaded, Shopkick seeks permission to use current location as recorded by the smartphone. In addition, Shopkick has a database of retailers and their geo-locations. It runs campaigns on behalf of the merchants and collects its revenues from merchants. Shopkick will let me know, for example, that the department store in my neighborhood would like me to visit the store. As a further incentive, Shopkick will deposit shopping points in my account for just visiting the store. As I walk through the store, Shopkick can use my current location in the smartphone to record my presence at the store and award points.

Jeff Jonas provided me tremendous motivation for playing with location data. I used *openpaths.cc*, a site that tracks cell phone location, to track my whereabouts for approximately three months. Watching my movements over these months was like having a video unfold my activities event by event. I could also see how I could improve the accuracy of the location data collected by openpaths with other known information such as street maps. With the help of a business directory, it is easy to find out the number and duration of my trips to Starbucks, Tokyo Joe's, and Sweet Tomato, my three most common eating hang outs.

Why would a customer "opt-in"? Device makers, CSPs, and retailers are beginning to offer a number of location-based services, in exchange for location "opt-in." For example, smartphones offer "find my phone" services, which can locate a phone. If the phone is lost, the last known location can be ascertained via a website. In exchange, the CSP or the device manufacturer may seek location data for product or service improvement. These location-based services could also be revenue generating. A CSP may decide to charge for a

configuration service that switches a smartphone to silent mode every time the subscriber enters the movie theater and switches back to normal ring tone once the subscriber leaves the movie theater. Prepaid wireless providers are engaging in location-based campaigns targeted at customers who are about to run out of prepaid minutes. These customers are the most likely to churn to a competitor and could easily continue with their current wireless provider if they were to be directed to a store that sells prepaid wireless cards.

These scenarios raise the obvious data privacy concern, which is a hotly debated topic worldwide. We will spend some time in the technical sections talking about data privacy, governance, and how consumer data can be protected and used only as permitted by the customer. As expected, there are many avenues for abuse of customer data, and data privacy must be engrained in the architecture for an effective protection of customer data.

3.6 Micro-Segmentation and Next Best Action

Automation has provided us with tremendous opportunity to use sensors to collect data in every step of the customer-facing processes, such as click streams in the use of a website. Sensor data gives us an opportunity to establish behavioral patterns using analytics. The early evolution was in use of analytics for segmentation. The original segmentations were demographic in nature and used hard consumer data, such as geography, age, gender, and ethnic characteristics to establish market segmentations. Marketers soon realized that behavioral traits were also important parameters to segment customers.

As our understanding grew, we saw more emphasis on micro-segments—specific niche markets based on analytics-driven parameters. For example, marketers started to differentiate innovators and early adopters from late adopters in their willingness to purchase new electronic gadgets. Customer experience data let us characterize innovators who were eager to share experiences early on and could be more tolerant of product defects.

In the mid-1990s, with automation in customer touch points and use of the Internet for customer self-service, marketing became more interested in personalization and 1:1 marketing. As Martha Rogers and Don Peppers point out in their book *The One to One Future*, "The basis for 1:1 marketing is share of customer, not just market share. Instead of selling as many products as possible over the next sales period to whomever will buy them, the goal of the 1:1 marketer is to sell one customer at a time as many products as possible over the lifetime of that customer's patronage. Mass marketers develop a product and try to find customers for that product. But 1:1 marketers develop a customer and try to find products for that customer."[16]

Early analytics systems were reporting systems that provided raw segmentation data to the marketing team so that they could use the data to decide on marketing activities, such as campaigns. Automation in marketing and operations gave us the opportunity to close the loop—to use analytics to collect effectiveness data to revise and improve campaigns. We are seeing surges in campaign activity. Marketers are interested in micro-campaigns that are designed specifically for a micro-segment or, in some cases, for specific customers. The customer experience information gives us criteria for including a customer in the campaign.

If a marketing analyst were to see my location data, they would immediately conclude that I travel frequently, both domestically as well as internationally. They could establish that when I am not traveling, I am typically working from home and occasionally at an office less than two miles from my house. They could also see a number of my regular activities. At Northeastern University in Boston, network physicists discovered just how predictable people could be by studying the travel routines of 100,000 European mobile-phone users. After analyzing more than 16 million records of call dates, times, and locations, the researchers determined that, when compiled, people's movements appeared to follow a mathematical pattern. The researchers stated that with enough information about past movements, they could forecast someone's future whereabouts with 93.6 percent accuracy.[17]

How do we use location data to derive micro-segments? At the simplest level, if we take the past three months of location data across a set of people, we can differentiate between globe trotters, people doing field jobs, "9 to 5ers" (i.e., people working desk jobs during regular office hours), and people working from home. At the next level, we can start to infer frequent behaviors. By observing how many times I visit a coffee shop, the mall, or a golf course, for example, we can establish my hang outs using frequency rules (e.g., "more than four visits per month, each for a duration of an hour or longer" constitutes a hang out). A marketer may seek a customer to "opt-in" their location information and offer location and context-specific promotions.

Next Best Action (NBA) recommends an activity based on the customer's latest experience with the product. This could include an up-sell or cross-sell based on current product ownership, usage level, and behavioral profile. An NBA could be offered any time the sales organization has the opportunity to connect with the customer via a touch point. NBA is far more effective in sales conversion compared with canned rules that repeatedly offer the same product over and over across a customer interaction channel. (Imagine your airline offering you a

discounted trip to your favorite warm-weather golf vacation spot on a cold day.) NBA can also be revised based on feedback from customer reactions.

Let me illustrate how Big Data is changing our business processes. For a number of decades, television producers relied on a control sample of audience viewing habits to gauge the popularity of their television shows. This data was collected using extensive surveys in the early days of television programming and then using special devices placed on a sample of television sets by companies such as Nielsen. With the advancement in the cable set-top box (STB) and digital network supporting the cable and satellite industries, we can now collect channel surfing data from all the STBs capable of providing this information. As a result, the size of data collected has grown considerably, providing us with finer insights not previously available. This information is very valuable because it can be used to correlate channel surfing with a number of micro-segmentation variables.

The grocery stores have been equally busy developing their understanding of customers. Most of them offer frequent shopper cards that can be used by the grocer to track purchase habits as well as used by the shoppers to redeem discounts and other useful campaigns. With identifying information collected from the customer, this shopper card can be correlated with a name and an address. So, if we have the retailer's information from the frequent shopper card and the cable provider's information about television viewing habits, we could correlate the channel surfing data with retail purchases by the household and insert appropriate commercials to run micro campaigns based on household purchases.

Retailers toyed with the idea of providing shopping gadgets to shoppers and eventually realized that creating a smartphone app to run on an existing device would be easier than engineering a new device. The shoppers may activate a mobile app as soon as they enter the retail store. The app starts to collect GPS-level accurate location information about the shopper and lets the shopper check in grocery items on the smartphone. At the checkout counter, the shopper connects the smartphone to the point-of-sale (PoS) device, and the grocery bill is automatically paid by the credit card associated with the app. As the person walks through the grocery store and checks in grocery items using a smartphone, a campaign management system starts downloading mobile coupons based on customer profile, past grocery purchases, and currently active promotions.

While advertising agencies have made the connection via messaging, we now have the ability to connect the dots at micro-segment or one-to-one marketing levels. That is, we can air commercials and see their impact on customer