Automating Intelligence

Industrializing Analytics at Enterprise Scale

By Thomas Casey, Consulting Director, Analytical Solutions

TERADATA

Table of Contents

List of Figures

TERADATA

Introduction

In recent years, a handful of high-profile organizations have brought to light the possibilities of using analytics to drive operational efficiencies, and this has resulted in explosive interest in the field of analytics. What was once thought of as a secondary support process (reporting and analytics) is now considered a strategic competitive asset. This evolution has driven some companies to develop new positions at both the practice level (e.g., Data Scientist) and executive level (e.g., Chief Analytics Officer).

Companies have more information about their customers than ever before. They collect and store petabytes of data and have a myriad of analytical technologies at their disposal. Despite an abundance of data and a strong desire to do something valuable with it, many companies still struggle to achieve a truly personalized relationship with their customers. Years into the analytic revolution, companies continue to focus mostly on technical platforms and find themselves often distracted from delivering business benefits. Given new technologies, use cases, and vendor solutions, why are most companies still leveraging analytics in rudimentary ways? Part of the answer is the level of complexity in understanding advanced analytics and associated concepts like Artificial Intelligence (AI) and how these can/should be applied to solve business problems. These concepts are not easy to fully understand, and vendors in this space sometimes demonstrate as much confusion in positioning their solutions as their would-be customers have in understanding them. It is always easier for organizations to look for the "right" technology versus integrating complementary technologies; however, leveraging analytics to optimize business processes requires the ability to look beyond the way things are done today—and focus on how they should work (which is not easy for most people).

The Changing Analytical Landscape

For years, there have been many prognosticators who talk about trends of the "connected customer." Personalization, 1:1 Marketing, etc., are not new concepts, and the goal of analytics to drive these initiatives has been in place in various forms for decades. Data warehousing and business intelligence solutions are well established, yet have remained relatively static for years in what they provided (albeit with incrementally new functionality) because these technologies were sufficient to deal with the analytical needs of a transaction-based world.

Pervasive and persistent connectivity has changed the paradigm from well-defined customer transactions to capricious and non-uniform interactions. This provides both new opportunities to directly or indirectly interact with people, as well as additional challenges in understanding why a customer does what they do. No longer do we simply look at an individual transaction to understand how a consumer communicates; even the view of omni channel is probably an outdated notion in this respect. Instead, successful businesses today enable analytically-driven processes by leveraging these interconnected touch points to communicate and continually learn about those they serve.

Ubiquitous access to broadband internet, additional devices with Wi-Fi capabilities, the growth in smartphone penetration worldwide, and the overall decreasing costs of each in relation to performance have created an environment where nearly any electronic device can be connected to the internet. In the future, one should assume that nearly every type of electronic device will be connected. These electronic devices can vary from the relatively simple (e.g., kitchen appliances, home thermostats, and personal devices) to extraordinarily complex (e.g., jet engine sensors). Conceptually, these are all electronic sensors that not only provide instant feedback on what is happening with respect to a device, but by being connected, can communicate with other connected devices. This concept is often referred to as the Internet of Things (IoT), and the opportunity to understand and ride the IoT wave is enormous. McKinsey estimates the potential economic impact of the IoT at up to 11.1 trillion dollars per year by 2025. By comparison, current global GDP is estimated to be 107 trillion dollars in Purchasing Power Parity terms.[1] Realizing these benefits will not be easy, as it will require a conceptual understanding of the use of analytics, the personalized nature with which they need to be applied, and the myriad of data hurdles that must be overcome to build out an effective and scalable solution.

Automating Intelligence—Enabling The Sentient Enterprise

In the book, "The Sentient Enterprise", the authors highlight five stages organizations undergo to achieve the objective where one can "position the enterprise in such a way that analytic algorithms are navigating circumstance and making the bulk of operational decisions without human help" 2. The introductory summary of this white paper series, "Operationalizing Analytics to Deliver Business Insights" outlined various trends driving this phenomenon and a framework referred to as Automating Intelligence was introduced. The framework is a pragmatic methodology for operationalizing analytics by looking at the processes, applications, and data architecture considerations necessary to build successful solutions in this domain. It is positioned as a journey and not all components need to fit into all use cases, but generally the framework provides a useful set of recommendations
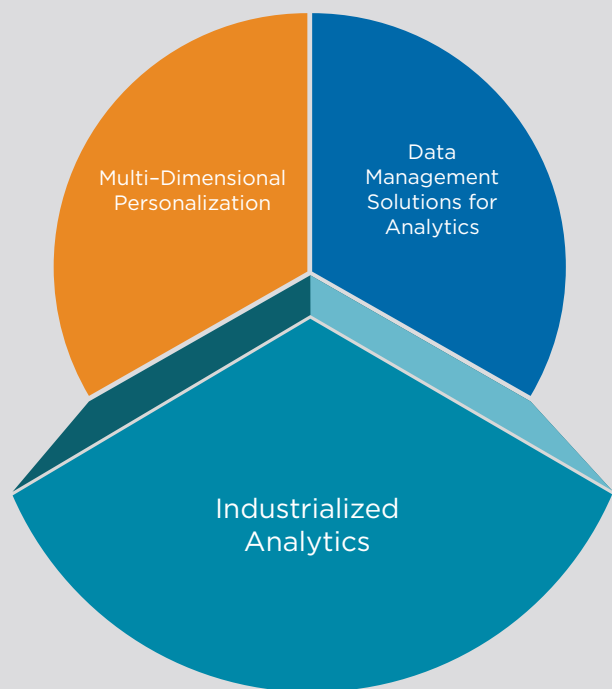
TERADATA.

Figure 1—Industrialized Analytics

This installment of the Automating Intelligence series deals with the first component of this framework, Industrialized Analytics. It does so by providing background on how the strategy was developed; the various analytical methods involved; how these methods map to an overall analytical process; and, finally, how these factors fit into the overall context of the Automating Intelligence framework. Where applicable, real-world examples outlining how each relevant concept works are provided.

This document is not intended to provide data scientists a discourse on or recommendations for specific statistical techniques. Instead, it focuses on giving business analysts, solution evaluators, operational management looking to leverage analytics, and executives developing analytically-led strategies a framework from which to apply and engage their discussions around how analytics should be operationalized within an organization. Other papers in this series include:

- *Automating Intelligence: The Real Facts About Artificial Intelligence*
- *Automating Intelligence: Taking a Multi-Dimensional Approach to Personalization*
- *Automating Intelligence: Recommended Strategies for Applying Recommendations*
- *Automating Intelligence: A Pragmatic Approach to Data Management Solutions for Analytics*
- *Automating Intelligence: Quality Data, the Cornerstone of Effective Analytics*
- *Automating Intelligence: Developing an Organizational Culture to Maximize your Analytical Initiatives*
- *Automating Intelligence: The Business-Led Strategy and Framework to Operationalize your Analytical Initiatives*

from which a series of guiding principles could be followed when trying to make analytics actionable. The framework is organized into three categories:

- **Industrialized Analytics**—the ability to take previous human interactions and use them to determine how current interactions should be defined and executed to encourage the necessary outcomes in an automated fashion
- **Multi-Dimensional Personalization**—strategies, techniques, and solutions for making non-human interactions personal
- **Data Management Solutions for Analytics** [3]—the architectural strategy to deal with all the data, analytical, and integration requirements necessary to operationalize analytics

---

1 http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/ The%20Internet%20of%20Things%20The%20 value%2of%2digitizing%20the%20physical%20world/Unlocking_the_potential_of_the_Internet_of_Things_Executive_summary.ashx

2 The Sentient Enterprise, The Evolution of Business Decision Making, Oliver Ratzesberger/Mohanbir Sawhney p 98

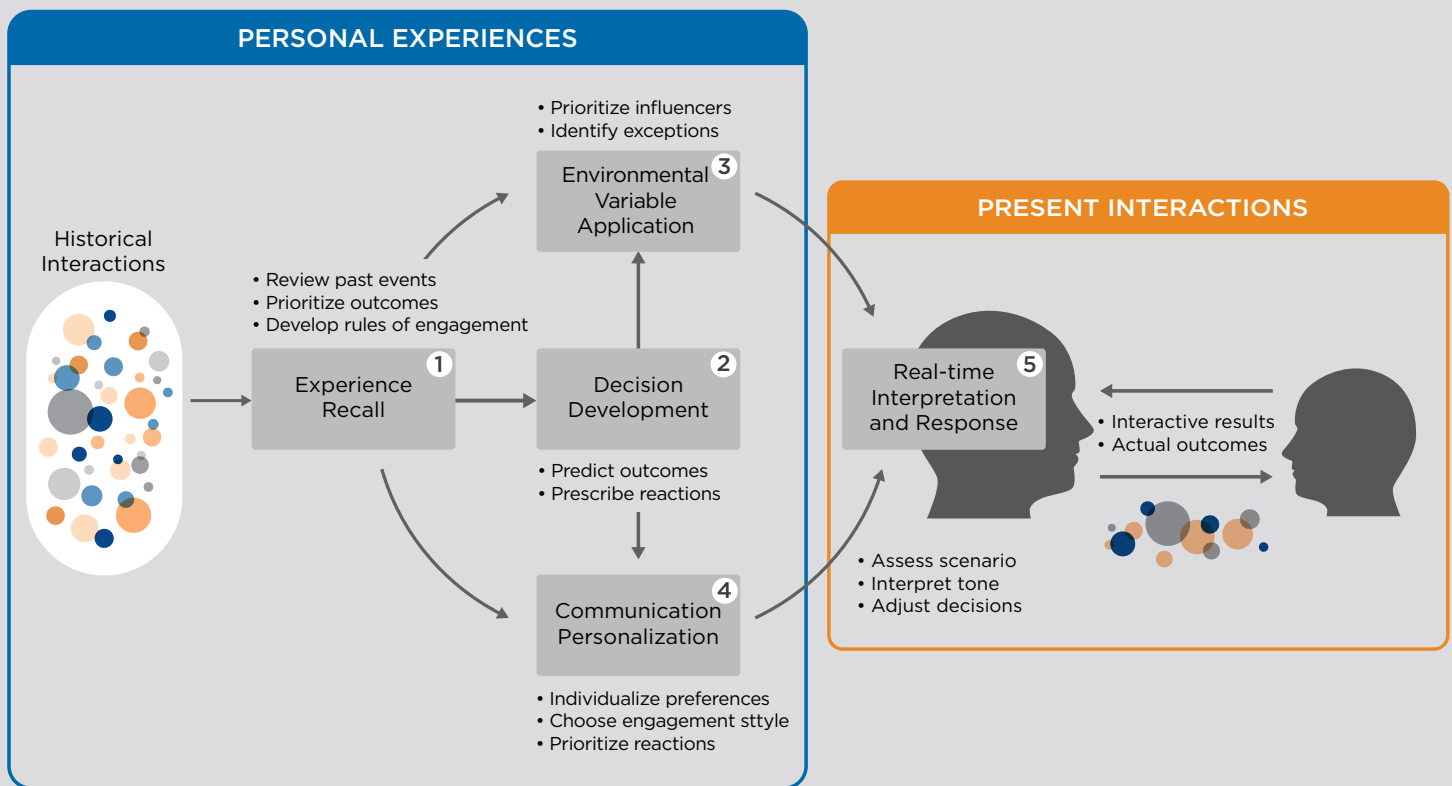3 Data Management Solutions for Analytics (DMSA) is a Gartner® term

TERADATA

Figure 2—Human Intelligence Defined

Industrializing Analytics by Emulating The Human Experience

Many people tend to grasp onto the latest analytical trends or concepts without understanding how the components work together. When it comes to the Industrialized Analytics portion of the framework and how it should work, it starts by first understanding the components of how human intelligence functions to ensure that solutions address each step in that process. This framework is based on the understanding that there are five independent components used to emulate the human experience. For a non-sentient process to evolve to this state, all five must be addressed, including:

- **Experience Recall**—What has one learned based upon past experiences? How have these experiences influenced behavior? How can they be applied in this situation? Much of what drives our reactions is based upon experience to drive to a desired result.

- **Decision Development**—Researchers have shown how past experiences are a primary driver in decision-making—especially when you must make complex decisions based on ambiguous information. Decision-making is informed by past experiences.[4]

- **Environmental Variable Application**—Everything that is known about an interaction may suggest a certain action, but there can be any number of implicit or explicit rules that can prioritize, adjust, or even override what is understood historically—especially when that is information that is only known at the time the interaction takes place.

- **Communication Personalization**—Every human has varying characteristics that drive the means and conditions by which we engage with them.

- **Real-Time Interpretation and Response**—Assessing a situation in real time to determine the specific scenarios for which our present interactions should apply. We might also look to interpret things like tone to serve as additional criteria from which to evaluate potential actions.

The previous four elements come together at this point to determine how we respond to someone. How well we align these five components determines the quality of the interaction.The way to visualize this is to think of how we make decisions in real time during interactions with others. Figure 2 outlines a high-level process on what influences our actions given the five components identified above.

---

4   https://www.sciencedaily.com/releases/2016/03/160311125019.htm
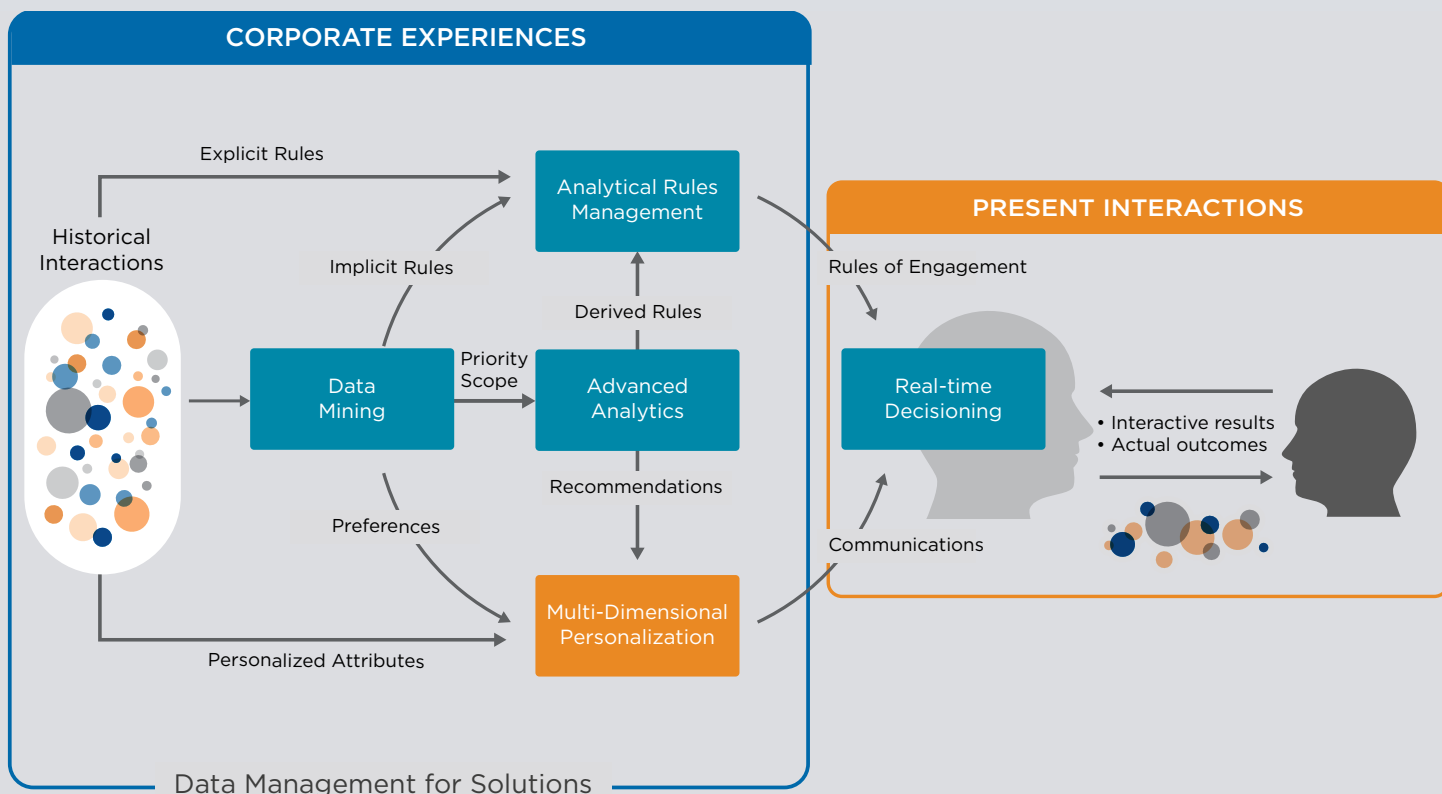
TERADATA

Figure 3—Autonomous Service through Operational Analytics

In addition to the five steps outlined, there is an unconscious process whereby results and outcomes are captured and retained for future use to enhance our ability to interact. This is important to consider as we look to build out any automated process that is designed to emulate intelligence.

Many of these analytical concepts may sound familiar, and conjure up specific thoughts on the manner and scope of each. The framework, however, takes a very precise definition of each concept and is specific in how they integrate as part of an overall process. The goal of Industrialized Analytics is not simply to apply previously defined concepts in an organized manner. Rather, it is to define a process using these functional categories to

Experience Recall	→	Data Mining
Decision Development	→	Advanced Analytics
Environmental Variable Application	→	Automated Rules Management
Communication Personalization	→	Multi-Dimensional Personalization
Real-Time Interpretation and Response	→	Real-Time Decisioning

replicate the way humans think and enable this process at whatever scale is necessary for success. The overall process and how it maps is illustrated in Figure 3. The details of the integration will follow in subsequent sections.

The teal blue boxes identify the steps required to automate analytics and deliver Industrialized Analytics at scale. The yellow box identifies the inflection point at which Multi-Dimensional Personalization and Industrialized Analytics intersect. The blue box represents the Data Management Solutions for Analytics category of the Framework. For the remainder of this document the focus will be primarily on the four teal blue boxes and the analytical techniques used to drive them. These four analytical functions are important and as they make up the other two legs of the framework, they are of equal importance in driving towards an overall solution.[5]

Automating The Analytical Process at Scale

Today there is enormous interest in developing and deploying autonomous analytical processes that will essentially improve themselves over time. One of the major trends discussed in the initial introduction to this series involved the significant advancements in AI.

5   These concepts are delved into detail in subsequent white papers created as part of this series

TERADATA

However, algorithms—even those that are self-learning—do not independently equate to automation. Analytics is used to provide the basis for the action (e.g., what kind of message should be delivered to this person under these circumstances); however, automation also requires some ability to understand inherent rules that may override the algorithm in specific circumstances. Deciding which takes precedence, algorithm or rule, requires context and the ability to make this decision in real time (with all available inputs). Furthermore, the channel may look at the original algorithm or rule input and decide that there is a better alternative based upon information that can only be known at the exact moment in which the interaction takes place. Combining these into a single process and doing so in a manner that is scalable is the fundamental goal of the Industrialized Analytics strategy and, as such, includes the following four components to drive this automation:

- **Data Mining**—the ability to analyze data to identify and prioritize use cases to be operationalized through analytics

- **Advanced Analytics**—using contextual data to make predictions or prescriptions using analytical techniques (which may or may not implement the use of self-learning technologies)

- **Automated Rules Management**—methods by which personalized rules are captured and applied to interactions as additional decision points

- **Real-Time Decisioning**—the ability for non-human initiated interactions to understand what is being requested in real time and to decide when to follow the analytical recommendation, leverage an interactive rule, or conclude that an entirely different result should be enacted

These steps can exist independently, but ultimately, they all need to be incorporated into a single set of actions to drive a truly interactive and operational analytical process; details of which are elaborated upon in the following sections.

Data Mining—The Compass of Our Analytical Journey

Many people believe that in the current world of analytics there is less of a need for traditional data mining activities. Some go as far as to say that automation of analytics through AI will make the whole concept of data mining obsolete. Machines may be good at applying algorithms, and even self-improvement of these algorithms. Nevertheless, the ability to provide context to the operating environment, qualify opportunities, define measurable descriptive variables, and prioritize efforts to apply predictive/prescriptive techniques is still something that is done through more traditional and human lead efforts—at least for now. There are several publications outlining what data mining is and the analytical processes it entails. For the sake of this framework, data mining is a process where:

Business transactions/interactions are grouped into business events,

These events are analyzed and evaluated to identify opportunities to improve the outcomes (often referred to as Descriptive Analytics),

Candidate business events are further analyzed to determine how they can be improved and the relevant effort (referred hereafter as Diagnostic Analytics), and

Efforts are prioritized to use analytics to predict what will happen (Predictive Analytics) and/or what should happen (Prescriptive Analytics).

Interactions

Business Events

Where do we see opportunities for improvement?
(Descriptive Analytics)

Potential business events we can/should use analytics to improve.
(Candidate Opportunites)

Why did it happen? Where can we make the biggest impact?
(Diagnostic Analytics)

Where should we focus our operational analytics activities?
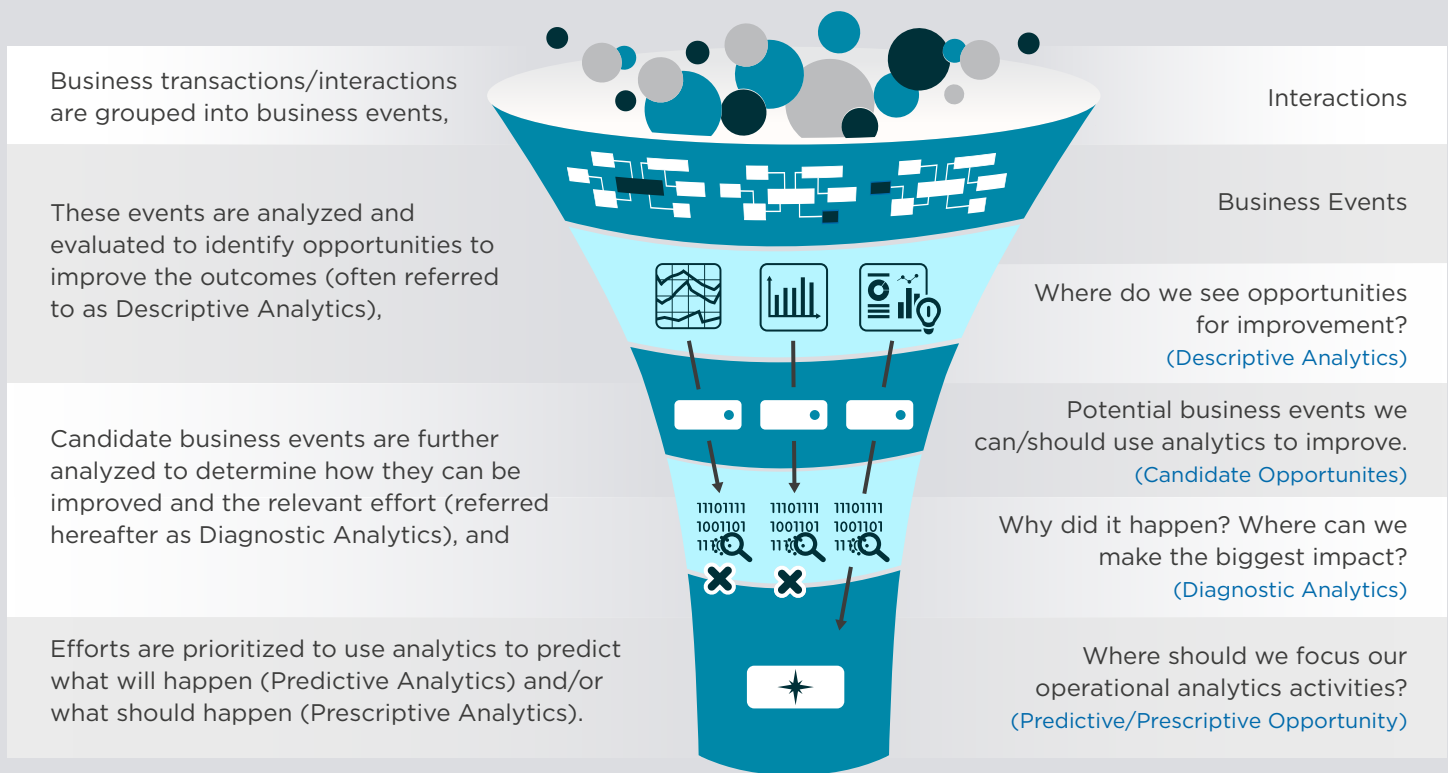(Predictive/Prescriptive Opportunity)

Figure 4—The Data Mining Funnel

Not all opportunities are equal and, even if an opportunity looks promising, not all opportunities can be addressed at the same time. Although Data Mining is often viewed as passé in this evolving world of analytics, this process is still very much a major part of operationalizing analytics because it identifies, prioritizes, and projects where to focus subsequent efforts with more advanced statistical processes. Although still very manual by nature, data mining should be viewed as a process and there are areas of opportunity for automation. "Slice and Dice" are not valid requirements and belie a lack of clarity as to what one is looking to do analytically.

This process ingests the transactions and interactions from traditional and IoT channels, which are then prioritized as depicted in Figure 4. This provides the direction for the efforts in advanced analytics activities, and may produce variables used for a statistical model. Additionally, data mining serves as a method for identifying implicit rules that can be used in guiding certain actions (see Analytical Rules Management).

These outputs can be delivered in a variety of formats, including segment definitions, clustering criteria, path to outcome definitions, sentiment drivers, or even just newly defined statistical variables than can serve as variables used in any modeling effort. Path to Outcome is based upon a series of events that determine the ideal path to achieve some desirable (or undesirable) result. Sentiment

analysis is the activity where interactive data (e.g., text, calls, etc.) is used to determine the feeling one has about a certain thing or outcome (e.g., customer satisfaction about a recent purchase). These can drive advanced analytical efforts that will then drive interactivity based on what we believe will or should happen.

Experimentation is a big part of the data mining process, and is a key component of scientific inquiry using analytics. The scientific method requires that one makes observations, questions the rationale for these observations, and ultimately formulates a hypothesis that will be tested in subsequent steps. There are two methods through which descriptive and diagnostic analytically inquiry occur: 1) Operational (using predefined analytics), and 2) Research and Development or R&D (using previously undefined analytics).

Operational methods involve analytical solutions that are standardized and deployed as part of a production-supported solution. Methods in this category have proven they need to be an ongoing, repeatable, and standardized method to perform some set of analytical activities (based on previous business validation). These solutions include analytical assets like standardized reports, dashboards, scorecards, exception reports, and threshold alerts. By standardizing these methods, a level of automation is possible with Data Mining tasks.
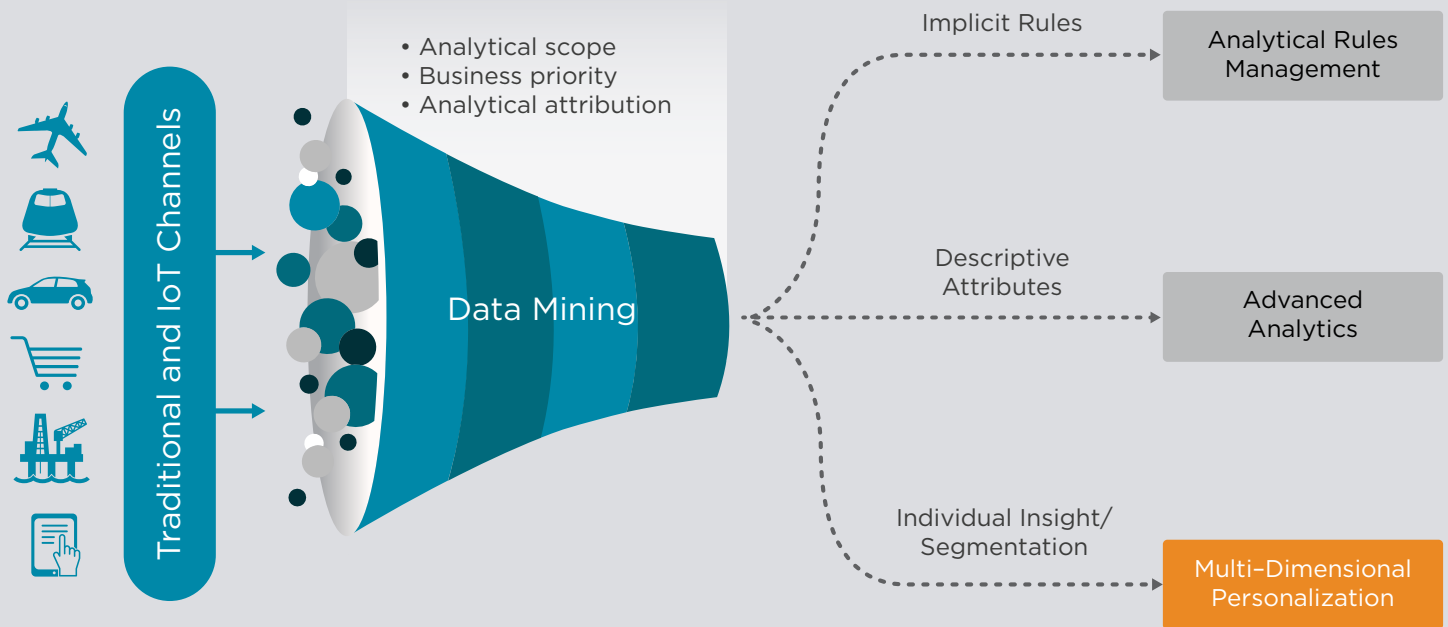
TERADATA

Figure 5—Data Mining Outputs

The R&D method allows for additional end user flexibility and autonomy to "test" preconceived notions through analytical rigor. Upon completion of any R&D initiative, there is a decision point to determine if the analytical outcome provides considerable enough insight or operational value that can drive ongoing analytical automation. If it does, there is value in having this as a repeatable and scheduled process whereby moving it from an R&D method to an Operational deployment. If not, then it is garnered insight and users should move on to the next analytical priority.

Keep in mind that most R&D initiatives should not lead to Operational deployments as most analytical efforts do not demonstrably improve an existing process or do not provide value as an ongoing analytic (i.e. it is a one-off action).

This is natural as part of any R&D type activity as it acts as a gatekeeper to enable users to explore ideas, but not force them all to be implemented. Nevertheless, any operationally deployed analytic should have first gone through the rigor of an R&D process.[6]

To better understand how data mining drives analytical automation, let us introduce Colleen. Colleen is a loyal customer for a large, luxury-brand retailer. This company prides itself on its intimate knowledge of its customers. It has physical locations, but also uses various digital sales channels and is active on various social media platforms. Based on previous analysis, this company developed a segmentation model based on lifetime value and has an

active program to pay special attention to its most valuable segment, to which Colleen belongs (descriptive analytics). Additionally, they have a series of mapped outcomes based on previous experiences that show customer triggers that lead to positive and negative outcomes (Diagnostic Analytics). Based on these measurements, this company has identified that Colleen made a recent purchase, but quickly returned the item with no follow up reshipment. Additionally, she has gone past her average time of purchase by 15 days, so is now part of a group of customers that require a series of proactive follow-ups based on her value and potentially negative perspective of the company. As part of this proactive program, Colleen has been added to an automated campaign that drives activities across any of the company's interactive channels using the insight gathered. Any specific insights that might drive decision-making preferences later in the process are identified as an interactive rule. Finally, this data (in a specified structure) is made available for a series of analytical processes to predict her future actions, and develop recommendations for next step actions.

Advanced Analytics—Using Algorithms to Predict and Prescribe

Making predictions and prescribing actions are the domain of most analytical discussions today. The advanced analytics component of the industrialized analytics process deals primarily in these two areas where statistical processes are designed to learn from historical observations that lead to better and more relevant interactions.

6   This process is explored further in AUTOMATING INTELLIGENCE—Developing an Organizational Culture to Maximize your Analytical Initiatives

TERADATA.

The primary method by which these techniques work revolves around input for real-time interactions based on historical interaction observations by making any number of predictions, prescriptions, and/or correlations. These serve as input to any decisioning interface by recommending the reactions to a given real-time interactive scenario.

Predictions are important because they help determine what we expect can happen, and do so before it happens. Common techniques are to predict the propensity to buy something or the potential for someone to dissolve the business relationship in a given timeframe. Some predictive techniques are not specific to the customer, but use customer specific elements to drive forecasts (e.g., forecasting the optimum size allocation for a clothing company in a specific geographic location based on the physical profile of customers in and around that area).

As for prescriptions, there are many different approaches to providing recommendations with varying levels of sophistication. This framework puts them into the following six categories:

- **Attribute Matching**—specific attributes that are shared across products are used to communicate product associations (e.g., suggesting a specific hose fitting when you buy an appliance, or offering leather cleaner when you purchase a new leather office chair)

- **Popularity**—determining the popularity of a product for a given set of customers and geographies (e.g., highlighting the top 10 products at a store you shop most frequently)

- **Item Association**—looking at individual transactions for all customers to determine the most statistically relevant relationships between products purchased together

- **Customer Profiling**—using demographics, behavioral, life stage, etc., to make associations in terms of recommended product (e.g., people with newborns getting offers for baby products)

- **Concept Mapping**—breaking each product into attributes and scoring these against a customer's profile/previous preference to make recommendations

- **Collaboration**—looking at filtering customers that look like you to make judgments about you (e.g., you look like Customer X, Customer X likes apples, therefore, you should like apples—even though you have never purchased apples)

Correlations can serve as inputs to predictions and/or prescriptions, as they provide the relationships between two things to determine if the correlation is positive, negative, or non–existing. Price elasticity is a classic example in that most people/product pairs exhibit a negative correlation between price and demand. As the price goes up, demand goes down. Nevertheless, there are some products (e.g., wines and alcohol) where people drive up their demand based on perceived value (based on the proxy, price) for some level of price increase. The strength of this correlation, either way, can be applied to make predictions and prescriptions for an individual.
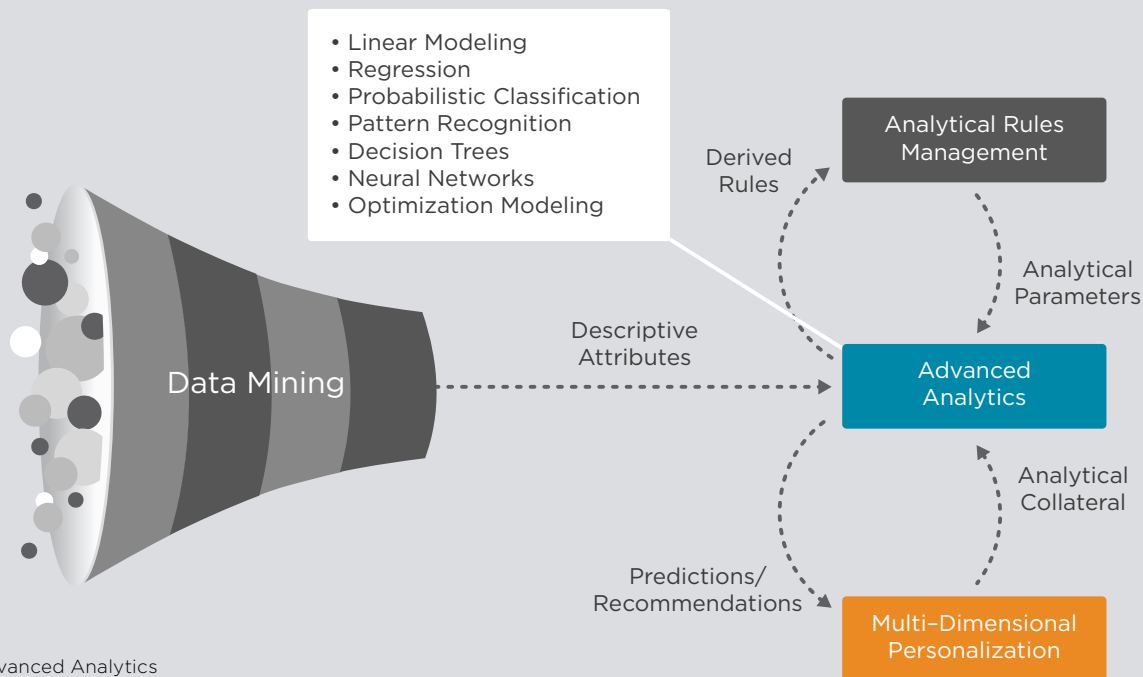
TERADATA

Figure 6—Advanced Analytics

There are other types of correlation that can stand on their own as either a prediction or prescription. One example is certain products that have strong correlations or affinities with one another and, as such, the strength of the correlation may drive a recommendation independently (e.g., people that bought Product X, also bought Product Y).

Using trends based on thousands of previous scenarios, it is determined that Colleen, who we determined is one of our top-tiered customers based on lifetime value, is showing signs indicative of churn. In fact, she is predicted to have a high likelihood of cancelling her company-branded credit card in the next 90 days based on a defined scoring algorithm (predictive analytics). Due to this prediction, she is recommended for a "white glove" set of strategies, which includes proactive engagement and special offers available for a limited time (prescriptive analytics). Looking at this as an opportunity to expand the share of wallet, these offers are explicitly targeting products she didn't buy but have the highest correlation, or affinity, to the top 10 items she has purchased in the past (statistical correlation).

Each of these concepts can leverage analytical techniques in various ways. There are several statistical techniques we equate to as "advanced analytics" with the following being some of the most popular:

- **Linear Modeling**—simple models with weighted variables to calculate simple predictions

- **Regression**—collection of more advanced modeling techniques that provide predictions that are not expressly linear

- **Probabilistic Classification**—produces a set of outcomes and the likelihood (i.e., probability) of an outcome being the right one

- **Pattern Recognition**—technique used to find patterns or regularities in a particular data set

- **Decision Trees**—series of event-driven observations that analytically drive to a conclusion that is used to predict future behavior

- **Neural Networks**—computational method whereby ambiguous patterns are deciphered by understanding what constitutes a known outcome, and applying that to determine if an unknown outcome is or is not equivalent to that known observation

- **Optimization Modeling**—many disciplines used to optimize a specific outcome based upon a series of analytical methods and rules

These techniques will drive predictions and/or prescriptions to be used in real-time decisioning. They can also serve by defining derived rules used to automate the analytical process (see following section for additional details). In addition to the outputs, the process shows that certain parameters (generated and managed as part of the Analytical Rules Management process) and analytical collateral (modeling objects used to define the scope of a prediction or recommendation—e.g. offers, events) are

inputs used by the Advanced Analytics process to help focus the modeling effort and results. Real-time decisions can provide feedback as the success or lack thereof of the modeling effort and help improve the algorithm over time.

Analytical Rules Management—Emulating The Expert

As stated in the discussion of advanced analytics, there are expert rules that can drive decisions and actions that may compliment or even override what the analytics suggest. Part of the industrializing analytics strategy requires we capture these rules and make them available to drive decisions. There are different categories of rules that drive this automation, which include the following:

- **Explicit Rules**—This requires an individual to self-identify rules of engagement for some type of interaction(s). One classic example is when a customer opts out of a specific marketing program, or sets a limit on how and/or how often they can be contacted. These rules typically supersede any other interactive action if there is a contradiction between this rule and some other prescribed action.

- **Implicit Rules**—These are rules that are not explicitly defined by an individual but rather deal with some insight gained indirectly based upon some knowledge of that person. These can be like explicit rules in that they may drive various aspects of the interactive process and can, in some instances, override other prescribed actions, but they are rules that were generated without an explicit request from the customer. An example of this type of rule is finding (through some analysis) that a person has an overall predilection to react more positively to emails where the background color is blue versus red. It does not mean the person consciously had to make that preference (that would then be an explicit rule), but because it was identified indirectly we can look for opportunities to use blue in our communications going forward.

- **Derived Rules**—These are rules where we learn, through analytics, something that can serve to address some level of process automation. An example is that we expect to see less than 5 percent of customers for each offer, but if we don't then the process should stop or that offer should be removed from the pool. Another example is that we have set a limit on how many people can receive a certain communication. Once that threshold is met, we no longer use it to drive additional interactions—even if the analytics tell us it is a suitable option for a specific individual. The biggest difference between an implicit and derived rule is that the implicit rule tells us something specific about the individual (derived indirectly), while a derived rule tells us how to deal with a group of customers when a certain condition exists (e.g., if anybody does X, then do Y).

These rules, as evident from the process, can come from various sources; including other processes. In addition, depending on the actual deployment and types of rules, some of these can be generated independently as part of the process itself. In addition to providing input to the real-time decisioning processes, this process can also establish parameters that are used as part of the advanced analytics process. These parameters may include thresholds (e.g., number of predictions per individual) or process drivers (e.g., run each algorithm on Monday at 5PM). The options are nearly endless, but there should be a central location where all rules are stored, reported on, managed, and validated. That is reflected in Figure 7.

In our example, Colleen explicitly stated that she did not want to be contacted via her smartphone, despite previous experiences that stated she reacted very well to smartphone alerts (explicit rule). Furthermore, through a series of data mining exploratory efforts, Colleen was shown to have a predilection towards animals, and additional insight suggests she has a pet even though she has never stated such. As such, whenever possible, content about pets and visuals showing people with pets are leveraged to drive personalized interest and potentially soften any adverse feelings she may have at present
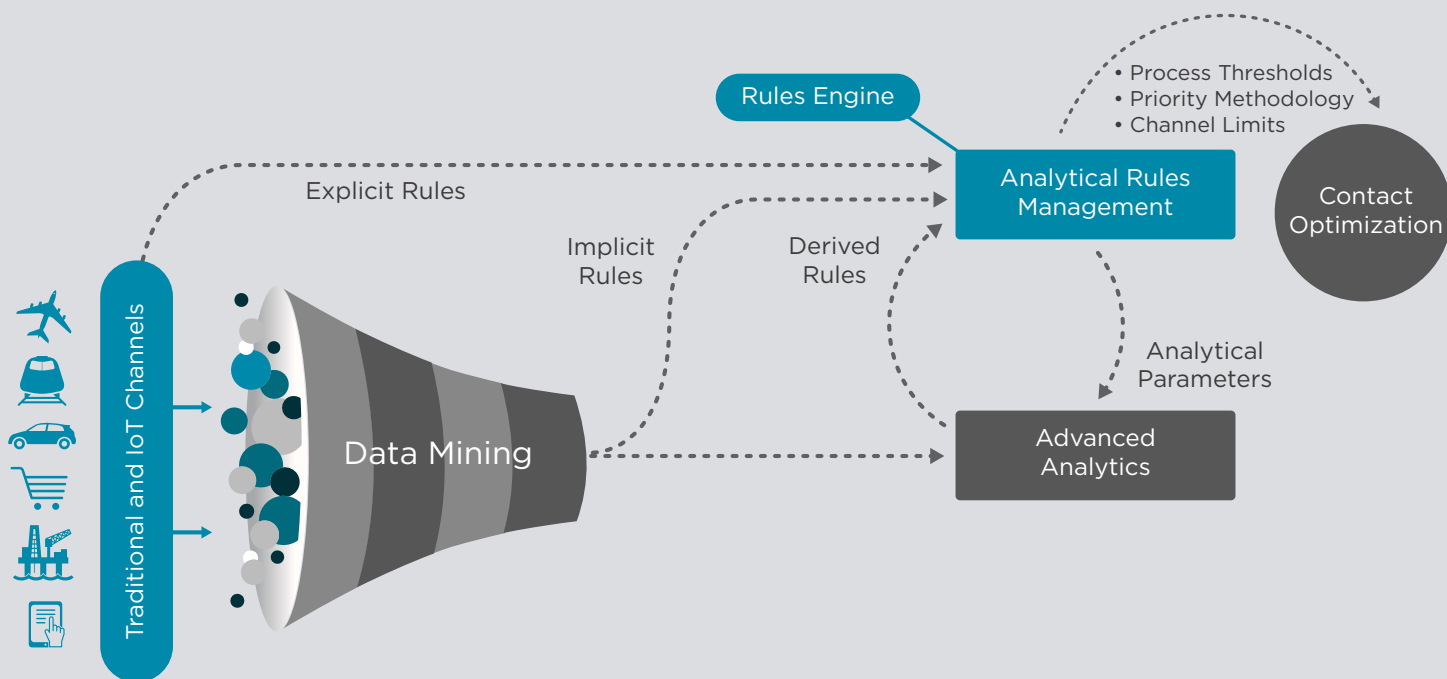
TERADATA

Figure 7—Analytical Rules Management

(implicit rule). Additionally, while looking to include Colleen in an upcoming campaign, it was noticed that she was a great candidate to highlight a cosmetic vendor with specialized messaging and content. Nevertheless, previous analysis has shown that customers exhibit a particularly adverse reaction to this vendor whenever they have a complaint about related products in the past 60 days (which she had). Given her situation with the company and this other insight, the system adjusts the recommendation to include the next best vendor to highlight to her—or anyone else in her situation (derived rule).

As shown, the analytical rules management process lends itself to supporting three different categories of rules that will drive analytical automation. One's solution, therefore, must cater to these rules whether they are derived from expert insight or analysis. Some rules can be generated independently from any single customer and some are customer specific. In either case, these rules and the inputs from the other processes are a key component in developing an optimized and autonomous process using analytics.

Real-time Decisioning vs. Real-time Analytic
Many people talk about real-time analytics, but what they typically are referring to is real-time decisioning. Most analytics have some time-variant applicability to their relevance so it stands to reason that the closer an algorithm gets to real time, the better its results. Although there is some truth to this in a few analytical use cases, the fact is that it is a broad misconception because, 1)

Most complex analytics are developed using very large sets of historical events and a few new interactions are not likely to demonstrably change the outcome, and 2) Real-time analytics denotes a use in a real-time interaction and is, therefore, restricted to something that can execute in web-based speed. This runs counter to many users' conceptions, but think about how a search engine works and the example becomes clearer.

We have all used search engines and we see they not only interpret our search request and bring back results in a split second, but also interpret what we are going to type and even "fix" mistakes by interpreting what we intended to type. When you break it down, the majority of what the search engine is doing has been formulated long before we ever conceived of our search. Web crawlers map the billions of internet pages to key search words and map this to the criteria in real time – but note, this is only doing a lookup in real time; the analytics have been developed well before any request is made. Subsequently, the interpretation of what we intended to say is another analytical technique that is defined and applied long before any request it made. It has some near real-time capability to adjust value based on recent, popular search requests. However, this again is not a case where the analytics are calculated on-the-fly. Finally, there are rules applied based on where a search happens geographically, or even if they want to move up a search result based on a sponsored link.

TERADATA

This, despite the ranking chosen analytically, is overridden because a rule (e.g., marketing spend) augments or even supersedes the algorithm's recommendation. So, the search engine solution certainly provides an example of real-time decisioning, but much of that leverages analytics that are anything but generated in real time. Our framework is similar in this respect and outlines real-time decisioning as having three interconnected yet discreet components: Contact Optimization, Interaction Management, and Interpretive Analytics.

- **Contact Optimization**—Manage both planned and event-driven communications using the following capabilities:
    - **Threshold Validation**—capping the number of customers for a single interaction or over multiple instances of that interaction
    - **Communication Prioritization**—ensuring individual relevancy by prioritizing interactions across a broader communication strategy
    - **Capacity Management**—ensuring an individual does not receive too many contacts (both within a single and across multiple interactions)
- **Interactive Management**—Involves all steps required for the proposed interaction to be validated and changed should a real-time element make the change necessary, including:

    - **Eligibility Validation**—Is the person still eligible between the time the baseline was created and now? If not, how does this change the ranking prescribed by the input analytics?
    - **Verifying Customer State**—Did anything about the person change (e.g., account status)? Should the interaction change completely or does it only alter a minor component of the interaction?
    - **Recommendation Modification**—Is there a specific set of conditions by which any previously determined recommendations must be re-prioritized or eliminated completely?
- **Interpretive Analytics**—Using analytical techniques, such as natural language processing (NLP), to interpret the nature of the interaction before the action is even completed. Includes the capability to assess effectiveness and provide feedback (supervised or unsupervised) to improve other advanced analytics algorithms.

Natural language processing (NLP) is a way to interpret real-time reactions, and can be used to better interact and adjust reactions based on the interaction interpretation. NLP may include one or more of the following techniques: 1) Automatic translation, 2) automatic summarization, 3) natural language generation, 4) speech processing,

---

7   The recommendations were not re-calculated in real time. The process just reprioritized the recommendations that were already made via the previous analytical process.
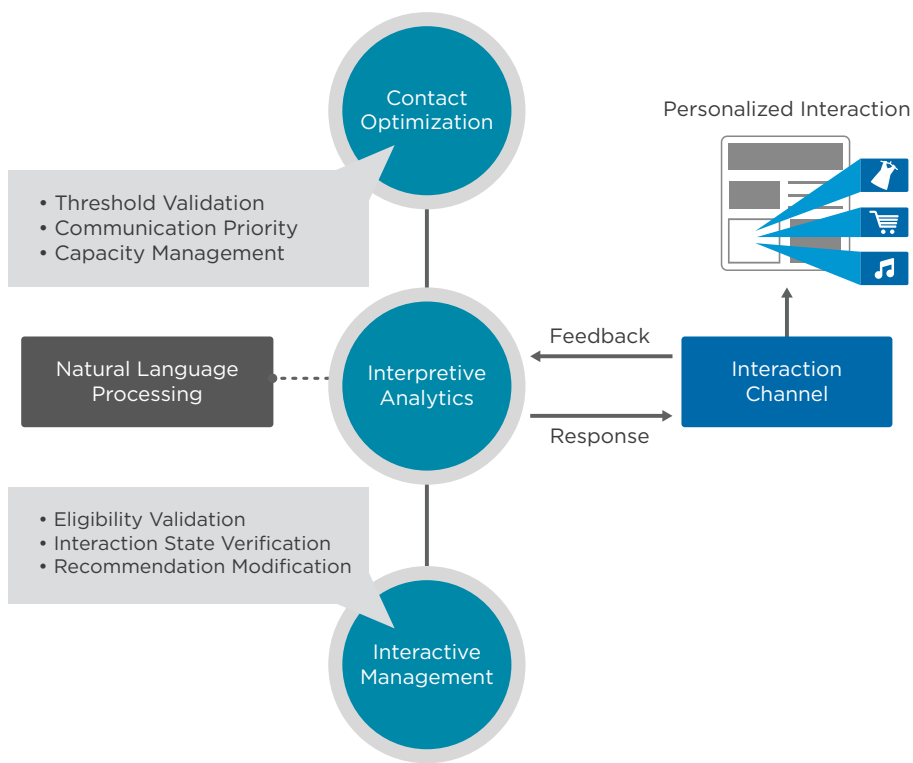
TERADATA

Contact Optimization

- Threshold Validation
- Communication Priority
- Capacity Management

Natural Language Processing

Interpretive Analytics

Feedback ← | → Response

Interaction Channel

Personalized Interaction

- Eligibility Validation
- Interaction State Verification
- Recommendation Modification

Interactive Management

Figure 8—Real-Time Decision Making

Leveraging these components in real time leads to the ability to emulate the human interactive process outlined previously. Decisions are made nearly instantaneously, but what goes into the decisions includes the ability to interpret situation, tone, predefined circumstantial rules, as well as experience (i.e., historical context). The combination of these components is shown in Figure 9.

Given how these inputs come together to develop real-time decisioning and ensure your analytical solution caters for these, a fully formed end-to-end, automated analytical process is possible

5) topic segmentation and information retrieval, 6) biomedical text mining, and 7) sentiment analysis. Many of these techniques are qualified use cases for deep-learning modeling techniques.

Colleen is interacting with an online agent in real time to address an issue with her recent store visit. Because of her situation and assigned program, her request was routed to the proper channel and prioritized based on her relative value (Contact Optimization). While on the phone, a series of follow up recommended next steps were available, but given the adverse nature of the interaction the recommendations were reprioritized in real time to reflect the state of the interaction and certain cues were reinforced based on her personal traits (interactive management). Finally, an analytical process that is designed to recognize and react to certain triggers in the interaction recognized a certain scenario and suggested something entirely based on where it interpreted this interaction was headed.

For instance, a popular item for Colleen is available from a local store (part of overstocked inventory) and to remediate the issue the agent offers to send it to her free of charge that day (Interpretive Analytics).

Thousands of these actions need to take place in a short period, and the responses for each need to be done in sub seconds. Therefore, the rules need to be clearly stated and staged such that the interacting channel can meet the necessary response requirements.

REAL–TIME DECISIONING

- Process Thresholds
- Priority Methodology
- Channel Limits

Analytical Rules Management

Contact Optimization

- Threshold Validation
- Communication Priority
- Capacity Management

Interpretative Analytics NLP

- Eligibility Validation
- Interaction State Verification
- Recommendation Modification

Multi-Dimensional Personalization

Interactive Management

- Analytical Ranking
- Content Assignment
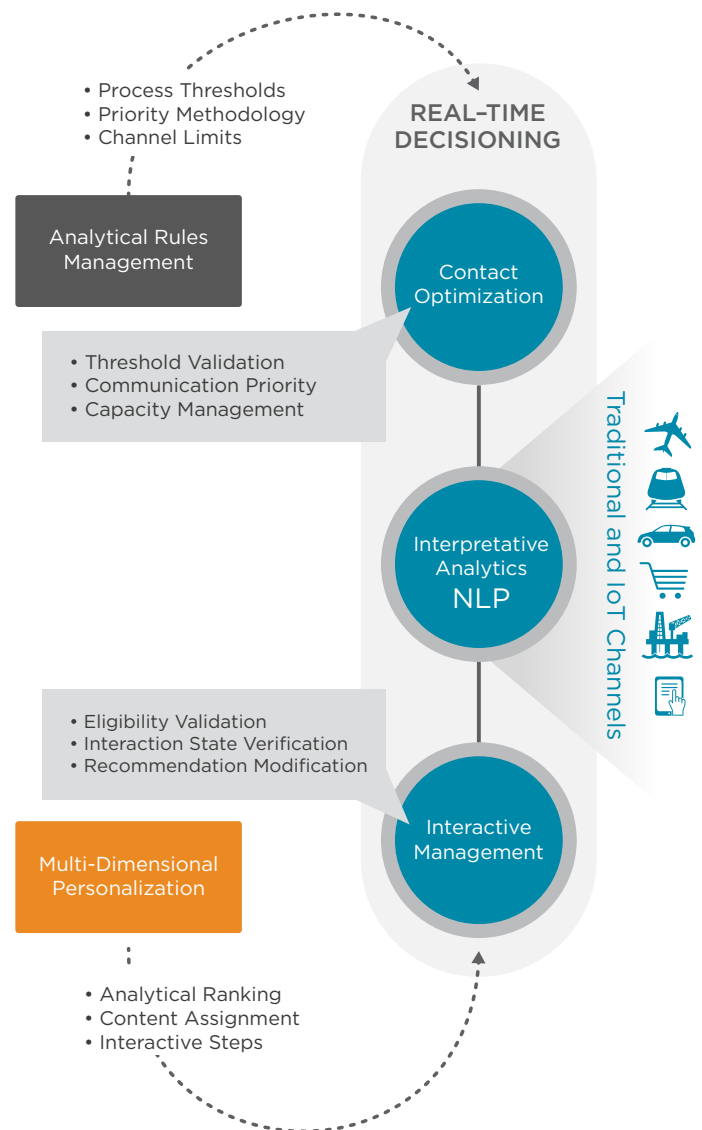- Interactive Steps

Traditional and IoT Channels

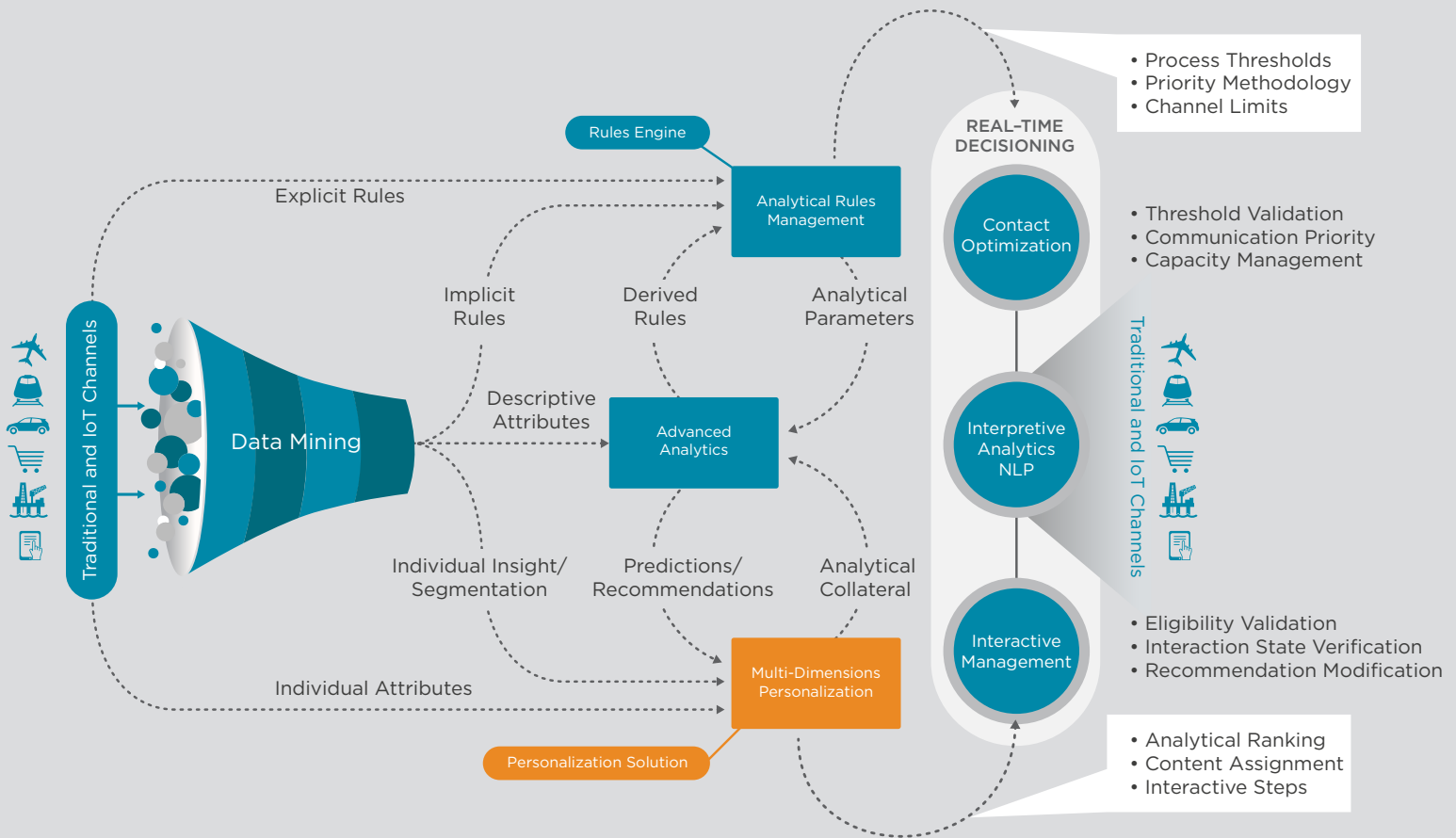Figure 9—Real-Time Decisioning Process

TERADATA

Figure 10—The Analytical Process

A Holistic View of The Analytical Process

The fundamental analytical process now comprises all the components necessary to automate and operationalize analytics in support of a particular interaction or set of interactions. This process provides context, insight, and the analytical building blocks to support the generation of more sophisticated rules, recommendations, and predictions. These analytics will in turn support real-time decisioning where different analytics and rules can drive the final decisions in real time. Finally, there is a mechanism by which a feedback loop is in place to allow the various analytics to "learn" over time and improve the predictions, prescriptions, and rules that drive the operational interaction. Finally, all but the initial setting of strategy and scope are set up to be automated along every step of the process. This end-to-end process is represented by Figure 10.

The key piece that is only alluded to in the above diagram relates to the second pillar of our Automating Intelligence solution, Multi-Dimensional Personalization. This includes the elements necessary to take a generic interaction and make it specific to the individual. This piece needs to be incorporated into the process but has its own set of conditions that drive how this is enacted. The following is an outline of the four-step process defined by the framework.

Step 1—Defining the Interactive Strategy: customer insights are leveraged to determine the basis from which interactions need to be personalized

Step 2—Personalization Setup: driving a personalization solution that is efficient and scalable by allowing various rules and optimization methods to drive the communication attribute values (e.g. content)

Step 3—Personalization Execution: creating a process with the means to execute these rules and update the attributes accordingly

Step 4—Channel Integration: requiring specific methods to integrate with the various interactive channel

The output of this personalization process is a set of personalized elements to include in the interaction as well as defined collateral that can serve as input variables to ongoing advanced analytical engagements. This is examined further in subsequent white papers but is presented here as a key representation of the integrated nature upon which the framework is designed.

In our example, Colleen was presented with an offer which was positioned based on her preferences and reflective of her importance and current situation. Internally, the offer was a best-case option for the business based on

TERADATA.

available goods in the location so that it minimized the give-away. Finally, the interaction was personalized to demonstrate intimate knowledge of her, did not include potential land mines that may have degraded the experience significantly, and ultimately re-established the good will between Colleen and the company. The only reason this was possible is that a holistic, integrated, analytically-driven, and automated process was in place to effectively manage the relationship. Keep in mind that this interactive scenario may take place across any interactive channel for thousands or even millions of customers every day. This is the power of industrialized analytics.

Guiding Principles for Industrializing Analytics

For analysis to truly provide benefit, it must be actionable. This capability not only denotes that the analytics provide insight into action, but also implies that the analytics should be as autonomous as possible to improve efficiency. Nevertheless, being autonomous is more than a self-learning algorithm; it requires a process to be defined such that it is scalable. This does not mean that the process does not increase as complexity increases. Instead, it means that the overhead required is predictable as dimensions are increased (e.g., double the number of predictions roughly doubles the amount of time to run the models). The solution should also be such that hard limits are not placed on the process (e.g., no more than 100,000 individuals scored at a time) or that any manual process (e.g., data mining) should have an open-ended time limit for completion.

Analytics always have a time-variant component to them, and their applicability changes as one moves further away from the time frame from which the analytics were generated. This does not mean all analytics need, or could, be rescored in real time, but it does state that the sooner it's applied and the outcome is expressed, the more applicable that original analytic will be in driving a desired outcome. Additionally, the industrialized analytics model expressly shows that there is not one model or technique that will satisfy all aspects of the process. Analytics drive different parts of the process, require different types of analytical techniques—or the combination of multiple techniques.

Finally, the automation model clearly defines an area that is most difficult to automate—data mining. This activity is important because it identifies opportunities outside

of the existing process and can prioritize these based on what is most important to the company, individual, etc. Still almost expressly a human endeavor, this can be a bottleneck in the overall automation process. There's a saying that if you give data scientists six weeks, they'll take six weeks; and if you give them six months, they'll take six months. So, whatever expectation one has for any step in the process, manual or automated, these expectations need to be set and managed accordingly. By expressly building it around an analytical culture of innovation with expressed expectations, it is possible to effectively execute these processes and provide meaningful direction, while minimizing the operational impact. Furthermore, whenever possible or applicable, making these activities operational will inherently drive automation.

Instead of looking at specific analytical techniques as a solution, treat analytics as a process enabler that will contribute to an overall solution. Different analytical approaches should be applied to different steps in the process. Subsequently, one should clearly define where these analytical components are used and where they act as inputs/output to other analytical processes so that a cohesive automation strategy can be designed from a holistic analytical perspective. Whether analytics are applied initially, ensure that all inputs are data-driven, not application-driven. This will allow you to apply automated capabilities in the future.

TERADATA

Where Artificial Intelligence (AI) Fits

Today, there is enormous interest in developing and deploying autonomous analytical processes that will essentially improve themselves over time. One of the major trends discussed in the initial introduction to this series involved the significant advancements around AI; however, AI is not a strategy. Machine learning is not a strategy. Both are important concepts—and one can be viewed as a subset of the other—but neither is a solution in and of itself. These are analytical techniques that may or may not be applicable to fit a specific need. Interestingly, many of the current AI techniques have been around for decades, but only recently has hardware progressed to the point that these techniques are capable of scaling on an enterprise level. Focus on the analytical process and apply these concepts if or when they make sense, and you have evolved to the point where you can make this judgment.

One of the more interesting aspects of AI and the concept of self-learning models is that this is actually a slight derivation to the long-standing process for model development. To understand this, one simply needs to look at the traditional statistical modeling process. Initially, these analytical models are probably developed and supervised to determine where improvements can

be made. Self-learning models for all predictive modeling techniques does not mean that people are completely removed from the analytical process. In addition to applying sound statistical practices to the initial modeling activities, the following guidelines should be considered with respect to self-learning algorithms.

- **Ensure business validity**—The model has no business context, it only uses the data it was designed to use as input, ensuring that the model and outcome make sense and accurately accounts for bias.

- **Focus more on data, less on complexity**—Models that appear to be statistically more significant but have less data to use in which to make predictions, very often perform worse than weaker models with more data. Having both is better, but if you must choose, pick more data.

- **Training and tuning is paramount**—Much like a student that still goes to school despite having learned a new concept, a model will need ongoing tuning and training to get better at making predictions.

- **Provide ongoing oversight**—Models will not be valid forever. You must continuously look for opportunities to evaluate, refine, and refresh them as required. This seems counterintuitive to the whole concept of self-learning algorithms but there are instances where environmental changes outside those used by the algorithm make the model obsolete.

With these techniques, a suitable set of activities describing the modeling process can be represented by the following four high-level steps:

- **Design**—the process used to identify a business need or optimization and the type of analytical approach(s) that will be used throughout the rest of the process.

- **Develop**—the process by which a data scientist develops a new model. Includes the development of different techniques and algorithms so that they can be tested against one another to determine the model that looks to be the best candidate to fit a specific business purchase.

- **Train**—the ongoing activities used to make an algorithm better (either supervised or unsupervised).

- **Evaluate**—when new external conditions exist outside of the initial model (or it produces values that are no longer deemed appropriate based upon evaluation criteria outside of the scope of the algorithm).

TERADATA.

Evaluate
On-going evaluation activities to determine if external factors have made this model obsolete?

Train
Use results & techniques to refine* the algorithm

*Refinement may be supervised or unsupervised (i.e. self learning)

Develop
Development of candidate algorithms and steps required to determine one champion

Design
The scope of the model as well as inputs required for development
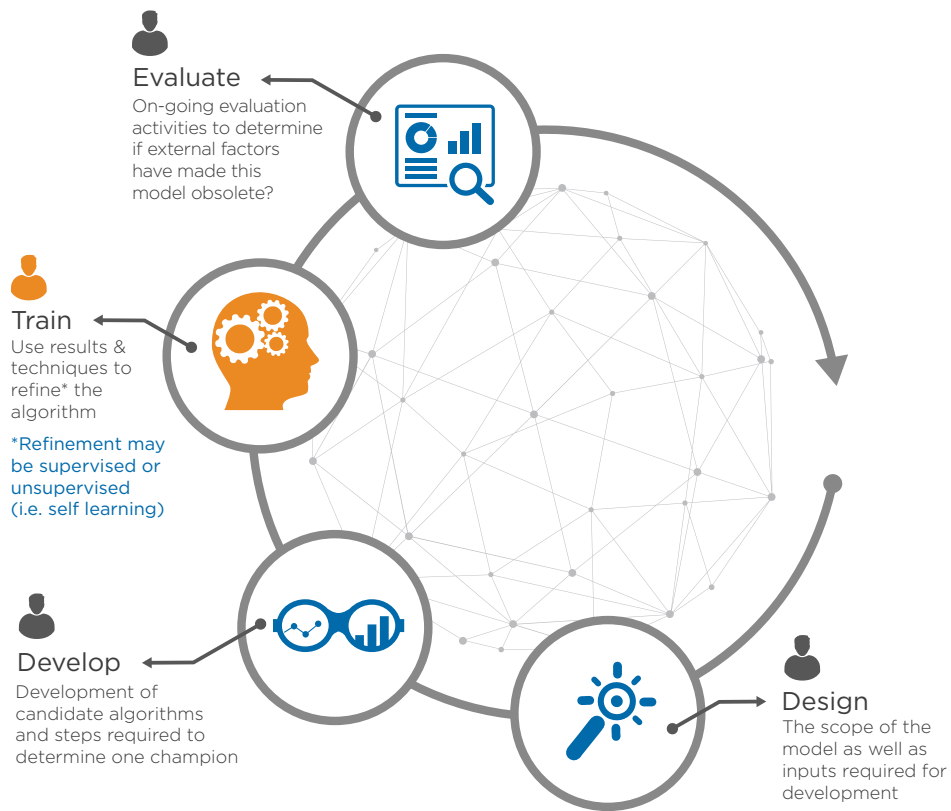
Figure 11—Advanced Analytics Modeling Process

In Figure 11, over time step 3 can be transitioned from supervised to unsupervised whereby inputs to the algorithm are automated and the testing/validation of results happens without direct involvement from a human being. This one step is essentially the conceptual approach to many (but not all) self-learning implementations.

Eventually, more advanced self-learning techniques can be applied to reach the goals of analytical automation; however, careful consideration should be applied to how this is implemented, which business processes it will enable, and how inevitable change management considerations will be managed. Machine learning does not come without risks.

With analytical autonomy, you potentially get results devoid of ethical considerations and altered perceptions, leading to a lack of trust in the outcome. As with the broader concept of AI, the area of self-learning statistical models should be entered with some caution and limited expectations. Much like the models themselves, users must learn how best to hand over analytical autonomy over time. The next white paper in this series will delve into further details on the phenomenon that is AI.[8]

8   AUTOMATING INTELLIGENCE—The Real Facts About Artificial Intelligence

03.18   EB10009