IBM SPSS predictive analytics: Optimizing decisions at the point of impact

Redguides
for Business Leaders

Theresa Morelli
Colin Shearer
Axel Buecker

- Enabling businesses, organizations, and governments to take the steps needed not only to survive but thrive in the current economic climate
- Using IBM SPSS Solutions to create greater value from data assets in all areas to deliver and act upon unique, predictive insights in real time
- Examining real-world industry applications of predictive analytics

Executive overview

Predictive analytics helps connect data to effective action by drawing reliable conclusions about current conditions and future events. It enables organizations to make predictions and then proactively act upon that insight to drive better business outcomes and achieve measurable competitive advantage.

In this IBM® Redguide™ publication, we explore IBM SPSS Predictive Analytics software both at a capability level and in the context of real-world application within various industries. We address the full suite of IBM SPSS Predictive Analytics software:

- IBM SPSS Data Collection for capturing attitudes, preferences, and feedback
- IBM SPSS Statistics Suite for research and analysis
- IBM SPSS Modeler for predicting future behavior
- IBM SPSS Text Analytics for mining unstructured data sources
- IBM SPSS Decision Management for optimizing operational decisions
- IBM SPSS Collaboration and Deployment Services for enterprise-wide management of analytical assets and results

This document includes examples of solutions using IBM SPSS Predictive Analytics, key concepts central to enterprise-wide adoption of predictive analytics, and architectural samples for proven predictive solutions.

A need for a new approach

Business leaders know that to meet their goals for profitability, revenue, cost reduction, and risk management, especially in the current economy, they cannot continue to operate the way they have in the past. Today's marketplace involves an exponential increase in the number and source of customer interactions; it is now a high-volume, multi-channel game. Organizations have to find new ways to compete and use technology to become smarter.

Through better management and use of information, business leaders can remove the blind spots that hinder informed decisions, and also achieve the next generation of efficiencies by providing precise, contextual analytics and insight at the point where these items can make a direct impact on business (*point of impact*). Doing so can enable *micro-optimization*, improving insight into patterns of customers, processes, and businesses, and deliver better real-time decisions and actions in every area of the organization.

This micro-optimization is made possible by establishing well-constructed processes and empowering individuals throughout the organization with *pervasive, predictive real-time analytics*. This approach can help shift from a *sense-and-respond* focus to a forward-looking *predict-and-act* focus. This approach also moves analysis from a back-office activity limited to a handful of experts to an approach that can empower everyone in the organization at the point of impact and in the context of the current situation. The result is rapid, informed, and confident decisions and actions throughout the organization, based on consistent and trusted information.

Analytics for a Smarter Planet

As a key element of enabling smarter decision-making through advanced analytics, SPSS was acquired by IBM in October 2009. SPSS had established itself as a leader in predictive analytics; SPSS drove the vision of predictive analytics and the emergence of this market. In its over 40-year history, though, the SPSS aim has been the same: *to drive the widespread use of data in decision-making.*

IBM SPSS Predictive Analytics (Figure 1 on page 3) is a software portfolio that provides data collection, statistics, predictive modeling, and deployment capabilities. These capabilities enable the end-goal of improving outcomes through decision management. IBM SPSS Predictive Analytics, collectively, is the combination of various capabilities that integrate multiple data sources for statistical, mathematical, and other algorithmic analyses, and decision management, which effectively deploys these predictions and allows actions to be taken based on these analytical insights. The results are higher quality decisions, measurably better outcomes, and higher returns on investment.

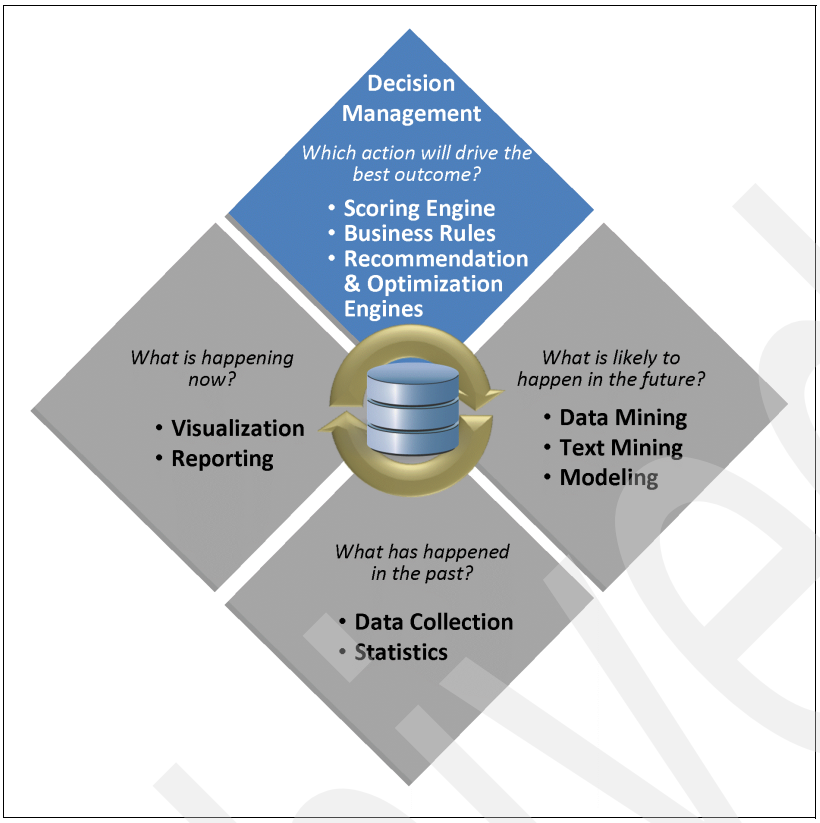

Figure 1 Predictive analytics overview

The evolution of decision-making

Historically, decisions have been made on the basis of anecdotal experience, or hunches, of seasoned domain experts. These “gut feel” decisions, though, are subjective and often inconsistent, thereby limiting their value.

Having the need to standardize key decisions and make them more consistent and reliable, many organizations have moved toward automated decision-making by using business rules. Although this automation provides a degree of efficiency and objective consistency, and improves the collective quality of decisions, static rules quickly obsolesce in ever-changing situations and conditions, and the limits of this approach become apparent.

Predictive decision-making, based on analysis of historical patterns and current conditions, is the basis for the highest quality means of making decisions. The reasons are because the models consider all available data, and also continuously adapt to new information, becoming smarter over time. With predictive analytics, the right decision for the given conditions can be made at the point of impact at the time when the decision needs to be made. Decisions are now customized for each unique case, rather than using generalizations for the aggregate. Figure 2 on page 4 depicts how decision-making continues to change.

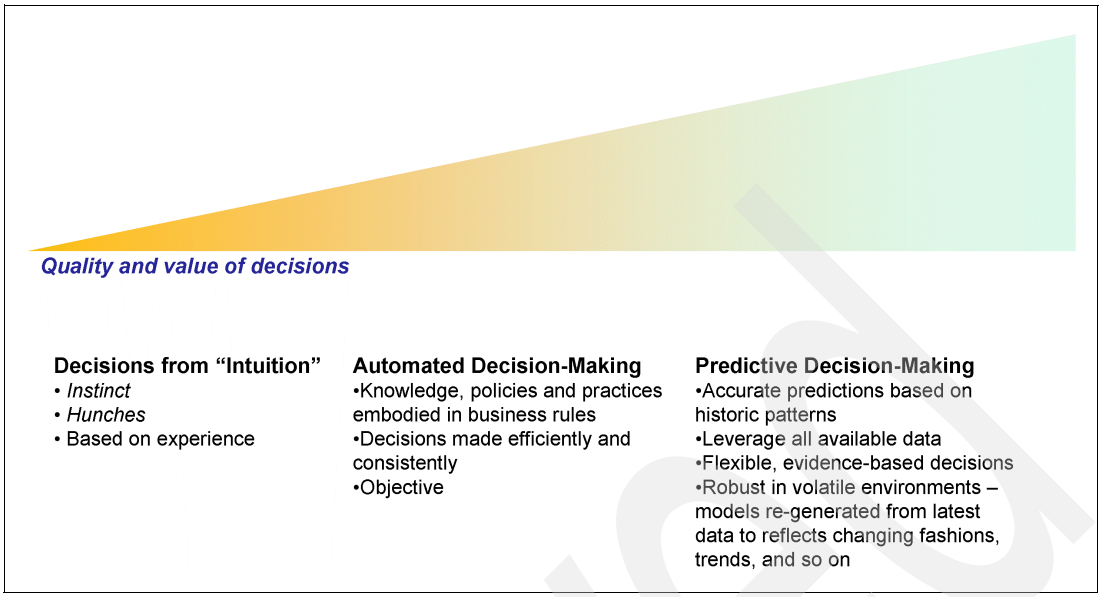

Figure 2 How decision making is changing

Why predictive analytics: The real ROI

The power of predictive analytics in driving optimal outcomes and profitable revenue growth is clearly demonstrated by organizations that deploy predictive solutions. An independent financial impact study by IDC[1] found that the median return on investment (ROI) for the projects that incorporated predictive technologies was 145%, compared with a median ROI of 89% for those projects that did not.

An independent assessment of SPSS customers found that 94% achieved a positive ROI with an average payback period of 10.7 months. Returns were achieved through reduced costs, increased productivity, increased employee and customer satisfaction, and greater visibility. Flexibility, performance, and price were all key factors in purchase decisions.[2]

IBM offers strong capabilities in information management, reporting and analysis, and with the addition of SPSS, now can offer users predictive power that leverages both structured and unstructured data[3]. This provides IBM SPSS users a distinct advantage as advanced analytics becomes a mainstream table stake in today's hyper-competitive marketplace.

[1] IDC Report: Predictive Analytics Yield Significant ROI - SPSS Inc., 2004, available at http://www.spss.hu/home_page/idcreport.htm

[2] Nucleus Research Report: The Real ROI of SPSS - SPSS Inc., 2005, available at http://www.spss.com/home_page/NucleusResearch.htm

[3] Nucleus Research Report: IBM and SPSS: Analytics for Everyone, 2009, available at http://www.spss.com/media/collateral/programs/IBM-and-SPSS-Nucleus.pdf

Current business challenges

This section describes typical business challenges that organizations face today in gaining actionable intelligence from their operations to optimize decisions and improve outcomes.

Universal challenges

Issues that are common to many organizations, regardless of industry include the ability to make high quality decisions, departmental silos throughout the organization and “rearview” thinking (for example, driving by looking only in the rearview mirror).

Decision-making

The ability to make high quality decisions is central to any organization. As the volume, variety, and velocity of information expands, this ability becomes increasingly difficult. With the state of current data explosion, decision-makers have to respond more quickly and with greater precision than ever.

As technology of all types becomes more robust and sophisticated, the data that is generated spans a multitude of formats, many of which are unstructured and machine-generated (for example, surveillance systems, GPS systems, and others). This situation forces a change in how data is managed so that it is meaningful in decision-making; organizations must look at ways to sort through the data, understand and connect the key pieces of information, and drive smarter business outcomes by leveraging the information assets.

Departmental silos

Many organizations develop natural silos based on various business functions, for example, finance marketing and sales, risk management, or operations. These departments often use separate systems to manage their business objectives, and when considered individually, they can realize a certain degree of effectiveness for driving various objectives. However, when information is not shared among departments, a significant opportunity to realize synergistic benefits is lost. For example, perhaps a marketing department is measuring target conversion rates for a particular campaign, but does not have visibility into which prospects might be bad credit risks and ultimately cost the organization money.

Although organizational, cultural, or political issues can be factors, this disconnect between departments is often a result of the existence of multiple disparate systems within an organization that articulate critical data in different ways: finance data is not directly meaningful to marketers; supply chain and assortment planning data is not expressed the same way as a customer profile. These various views and representations between systems perpetuates the silos and inhibits cross-departmental intelligence. To mitigate this issue, data of all types from multiple sources and systems must be integrated into a common vocabulary that is universally meaningful.

Rearview thinking

Traditional forecasting and business intelligence is grounded on historical data, where expected future outcomes are based on what has happened in the past. This method would be reliable if all conditions and variables of a given experience remain absolutely static, which is clearly impossible given the dynamic nature of people, economies, environmental conditions, technologies, and other variables that affect every decision point in an organization’s operation.

Historical data does become critical in decision-making when it is combined with current data and continuously refreshed with new information. This approach is the basis of predictive

decision-making; models are based on what has happened and why, and also on what is happening now and what will likely happen next.

Limitations of traditional approaches

This section describes the ways in which traditional approaches are limited, including types of data used, inability to personalize communications, and a general need to move beyond traditional ways that analytics are positioned within an organization.

Disparate and incomplete data

Many traditional systems, such as those that generate customer interactions, forecast inventory needs, or support public safety, often rely on broad segmentation and historical data that generalize, rather than personalize, customers and constituents. These forms of historical data may include demographic data and transactional history and are often expressed in structured formats that allow for fundamental analysis (*slicing*). However, because behaviors are motivated by a variety of dynamic human factors that extend beyond static datapoints, relying on such data provides an incomplete picture. This is particularly true when considering the rich insights that are available in many unstructured formats that are overlooked in fundamental analytic solutions. These data types, which comprise the vast majority of available data, provide critical insights into motivations behind various behaviors such as preferences and desires of customers.

This attitudinal data is the key to differentiating customers within a larger segment. For example, individuals living in a particular postal code might differ significantly in their reasons for living there (convenience, proximity to job and family, culture and weather, and so on). Understanding the underlying reason for a particular behavior unlocks the insights in otherwise static datapoints that allow for prediction of future behaviors.

Similarly, visibility into performance and metrics using business intelligence (BI) tools is critical to managing performance, as is the ability to perform financial analysis. However, if the data is incomplete (that is, limited to historical instances only), the ability to make high-quality decisions throughout the enterprise becomes severely limited.

One-size-fits-all and one-way communication

The systems through which an organization interacts with “the world,” for example, with its customers or constituents, often convey single, undifferentiated messages and usually support only a one-way flow of information. This approach misses opportunities to tailor communication and to learn from interactions.

Standard customer relationship management (CRM) systems, for example, are fundamental to managing customer interactions. They ensure that interactions occur consistently (often through automation) and efficiently, and are recorded. However, these CRM systems also have severe limitations with regard to understanding individual customers and creating customized interactions that will ultimately be more effective and profitable. They often stop at the *one-to-many* level rather than offering the finer-grained differentiation that creates a *one-to-one* customized experience. Traditional systems are also often unidirectional: they “shout” offers at customers with little attention to personalization.

The need for a paradigm shift

True predictive capability is absent, or at best rudimentary, in traditional industry solutions. As a result, organizations focus on sensing and then responding to issues, which is a reactive stance. Often, the damage (churn, fraudulent activity, excessive marketing spend) has already occurred, and the goal then becomes to minimize the impact on the organization. Traditional organizations also often rely on the instinct and intuition of resident experts rather

than using objective evidence that is grounded in data to guide decision-making despite the empirical evidence showing that statistical models consistently *outpredict* the experts. Many analytical systems also often require specialized skills or training, limiting their use to a select number of individuals within an organization and constraining the analytic activity to back-office functions rather than positioning it at the critical point of impact.Automation is also widely touted in traditional systems. Although the ability to automate is important, even more important is the optimization of decisions, or making the *right* decision at the right moment (Figure 3).

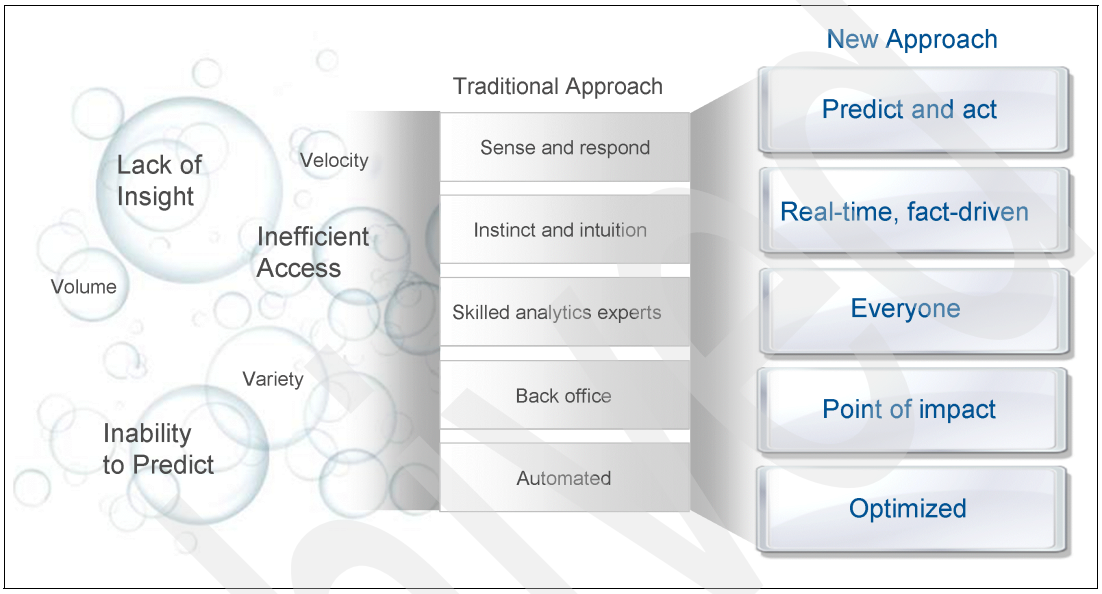

Figure 3 Addressing new business challenges with business analytics

Sector considerations

This section has information about perspectives and operational objectives of the private and public sectors.

Private sector

Most commercial organizations share similar goals: attract the best, most profitable customers; grow the business through cross- and up-sell; retain high quality customers; minimize risk to the organization; and detect and prevent fraudulent activity. Supply-chain considerations, such as effective assortment planning in retail stores and efficient production lines in manufacturing, are also key considerations that can significantly affect the customer experience and ultimately affect customer value. In short, the goal is to grow revenues and minimize costs, thereby producing profitable revenue growth.

Typical application areas are as follows:

- Attracting the best, most profitable customers through well-targeted campaigns.
- Increasing revenues through cross- and up-sell to new and existing customers.
- Reducing defection of high-quality customers and, conversely, identifying those who are costly and should be allowed to go through attrition.
- Minimizing the effect of fraudulent activity by focusing the work of investigators appropriately.
- Increasing customer satisfaction through faster response and processing of legitimate claims.

- Building customer loyalty through effective and reliable inventory management.
- Reducing operating costs by predicting maintenance needs proactively.

Public sector

Public agencies manage a unique set of challenges and often face additional scrutiny because of the personal nature of public security, healthcare, and education. Because they are also often funded or subsidized by government-run agencies, these organizations face additional pressures and bureaucracy around operational budgets.

Typical public sector application areas are as follows:

- Government agencies manage functions as diverse as tax audit selections, military force recruitment, and proactive policing and public safety.
- Healthcare organizations seek to proactively manage their resources and fine-tune their practices to provide better patient care.
- Colleges and universities manage the entire student life cycle more efficiently, recruiting the right mix of students, offering students a selection of programs and assistance to keep them enrolled, and managing alumni development programs with greater success.

Industry examples

This section has information about functional requirements and challenges of specific industries, and describes predictive analytics solutions to address the gaps that currently exist. Industries include consumer banking, insurance, telecommunications, retail, distribution, industrial, and the public sector.

Banking

This section describes the business and functional requirements of the banking industry, and offers examples of how IBM SPSS Predictive Analytics can address those industry-specific needs.

Business and functional requirements

Banks today are faced with a low or declining share of wallet, rising marketing costs because of low response rates and unprofitable campaigns, and a general inability to effectively segment and target customers. Banks struggle to understand the needs of individual customers and therefore often manage interactions and offers at the aggregate level. As a result, campaigns are poorly targeted and marketing costs often exceed budgets, while lead generation goals still are not met. Within operations, banks often implement multiple systems for lead-sharing, reporting solutions, and a metric tracking mechanism, which ultimately prevents effective and accurate information-sharing throughout the enterprise.

Banks are also under increasing pressure to minimize risk and focus on customers and prospects that will be profitable, rather than result in costly liabilities. Although banks typically have large, multiple data warehouses for risk and compliance, complex and inflexible risk-modeling methodologies prevent effective use of this information. The ability to predict who will be high risk is critical, driving the need for a platform that synthesizes disparate risk and finance data, and predicts which customers will be risky, and with whom they should not do business.

Information technology (IT) executives within the banking sector typically look to maximize returns from investments in warehousing and CRM, and find new and innovative approaches for supporting business goals. Within the lines of business (LOB), departments look to better understand buying preferences, future financial needs, and transaction behaviors of their customers to be able to provide high-quality service that is consistent with those needs.

Key focus areas: Marketing optimization (targeting), cross and up-sell, risk mitigation

Predictive analytics approach

To address the key needs of commercial banks, IBM offers multiple solutions that span marketing goals, risk mitigation, and a combined solution that integrates both objectives.

IBM SPSS Campaign Insight and Offer Optimization capability uses dynamic algorithms that use both structured and unstructured (textual) data to identify high-quality leads that are most likely to respond, allowing for more effectively targeted campaigns and higher returns with less marketing money spent.

Risk analysis can also be integrated for a more robust solution, which can ensure that the business driven by marketing offers is truly profitable and does not increase risk exposure. Accurately assessing risk using predictive modeling instead of rigid business rules can prevent potentially profitable customers from being filtered out based on blanket risk scoring. *IBM SPSS Integrated Risk Management* solutions can reveal factors behind individual risk

scores, empowering banks to refine business rules and personalize the handling of each customer to capture opportunities that might otherwise have been lost.

A bank may use an outbound marketing approach, using a campaign management system to deliver offers to customers through direct mail and track campaign effectiveness. Customer lists are selected based on propensity to respond (as dictated by the output of the predictive models), then filtered to remove prospects who are a credit risk. The filtered lists are passed to the campaign management system, which executes the campaigns and delivers the offers.Saving the purchase propensity score and the credit risk score to the customer data warehouse also enables using the combined scores to subsequently control offers of this product through other channels, for example, when customers log on to online banking or use an ATM.

The propensity models are built from historical campaign records of people who responded (or did not respond) to offers for similar or identical products. The credit risk models may use externally sourced risk scores as one of their inputs, but typically refine those "one-size-fits-all" scores by adding many other variables and looking at the history of which customers proved to be good or bad risks in this business. The output of these models (scores) are then combined, at which point human expertise can be introduced to manage and resolve combined scores that work in opposite directions. Although thresholds may be set to veto all offers when the associated risk exceeds a specified limit, lower bands of risk may be accepted under certain conditions, for example, if the propensity score and predicted profit are both sufficiently high and the predicted amount of risk exposure is below a specified limit. This solution is depicted in Figure 4.

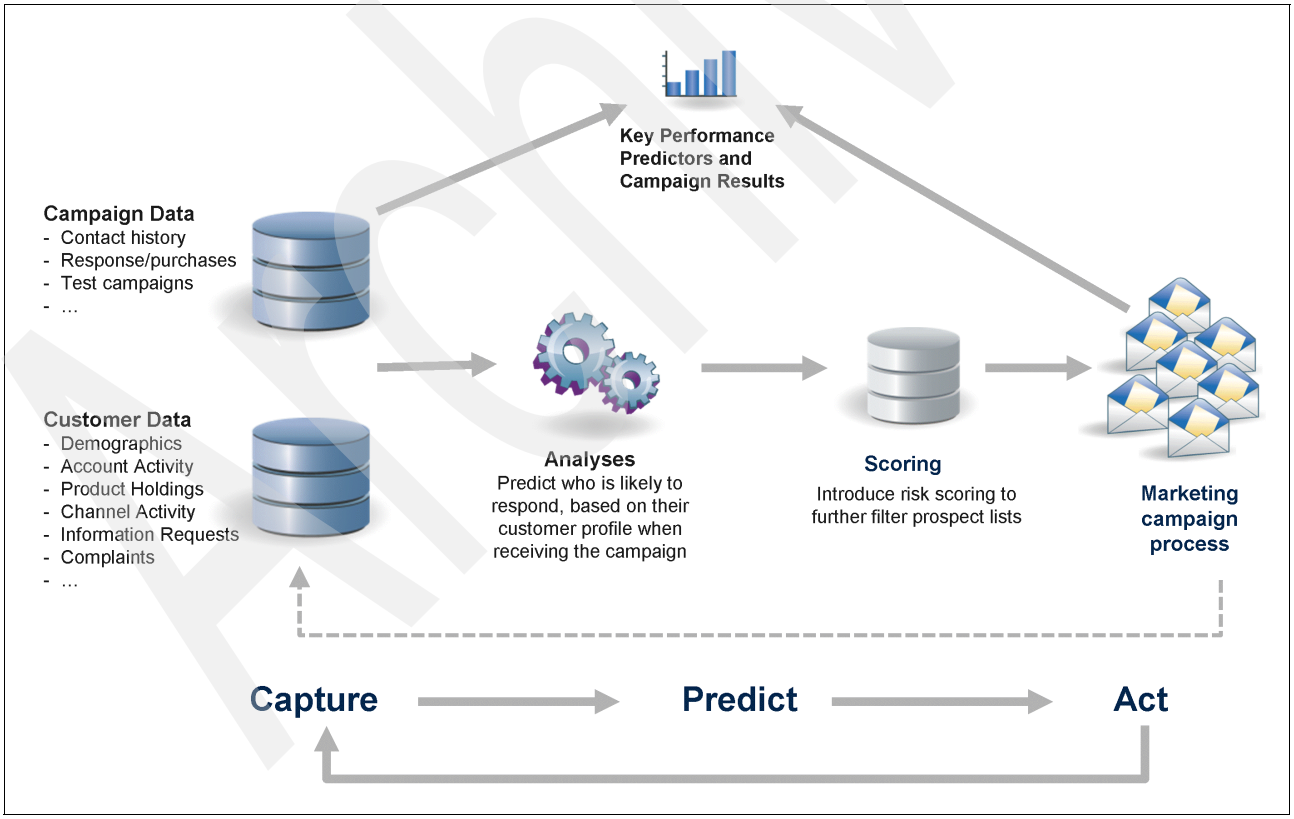

Figure 4 Analytical process, integrated banking solution

Insurance

This section describes the business and functional requirements of the insurance industry, and offers examples of how IBM SPSS Predictive Analytics can help address those industry-specific needs.

Business and functional requirements

In Insurance, churn levels can be as high as 30%. One primary goal of this industry, therefore, is to reduce the churn rates of the most profitable customers while minimizing the cost to service less profitable customers, such as migrating them to more cost-effective channels. Because insurance spans many types of products (auto, homeowners, health and life, and so on), effectively targeting prospects and existing customers with the right product mix to meet individual needs is crucial to driving revenue. Because of the time sensitivity in preempting churn, insurance organizations must be able to act timely and decisively at critical *moments of truth* in the customer experience.

Claims handling is one of the most critical business processes for insurers. There is obviously a need to identify and avoid paying fraudulent claims, but this has to be balanced against the cost of deeply scrutinizing every claim in case it is fraudulent. Although the approach that many insurers take involves going through a significant level of information-gathering and verification on every claim, this approach incurs costs, even for claims that reveal themselves to be perfectly legitimate. The approach also leads to what can be a lengthy process, often leaving "good" customers dissatisfied because of delays in reaching settlement and payment, and an impression that they are under suspicion as they are repeatedly asked for further information. Additionally, identification of claims to be passed to the investigation team tends to be on the "gut feel" of claims handlers. As a consequence, many claims passed to the investigators are *false positives*, consuming their time and effort, and making them less efficient and less successful in detecting genuinely fraudulent cases.

IT stakeholders are typically interested in summary-level reporting and planning technology with greater visibility into information and requiring fewer IT resources. Because of the broad availability of web-based information (for example, online comparison of insurance products of competitors) customer self-service capabilities are also critical. LOB managers are generally interested in improving accuracy of sales forecasting, consolidated and higher quality customer profile information, and increased marketing conversion rates per dollar spent.

Key focus areas: Churn reduction, detecting and preventing fraud, marketing optimization for cross-sell and up-sell

Predictive analytics approach

IBM SPSS Customer Retention and Growth solutions for insurance use advanced algorithms to identify at-risk customers and recommended retention strategies, then strengthen customer relationships by analyzing attitudes, needs, and preferences of policyholders to create customized offers that drive customer loyalty and sustained premium income.

This process is similar to the banking solutions, involving propensity scoring for outbound campaign to ensure that the right offer is extended to the right customer through the optimal channel. Similar processes are used in prospecting, and also in increasing customer satisfaction (and therefore loyalty and retention) of the most desirable customers. The cross-sell and retention solution is depicted in Figure 5 on page 12.

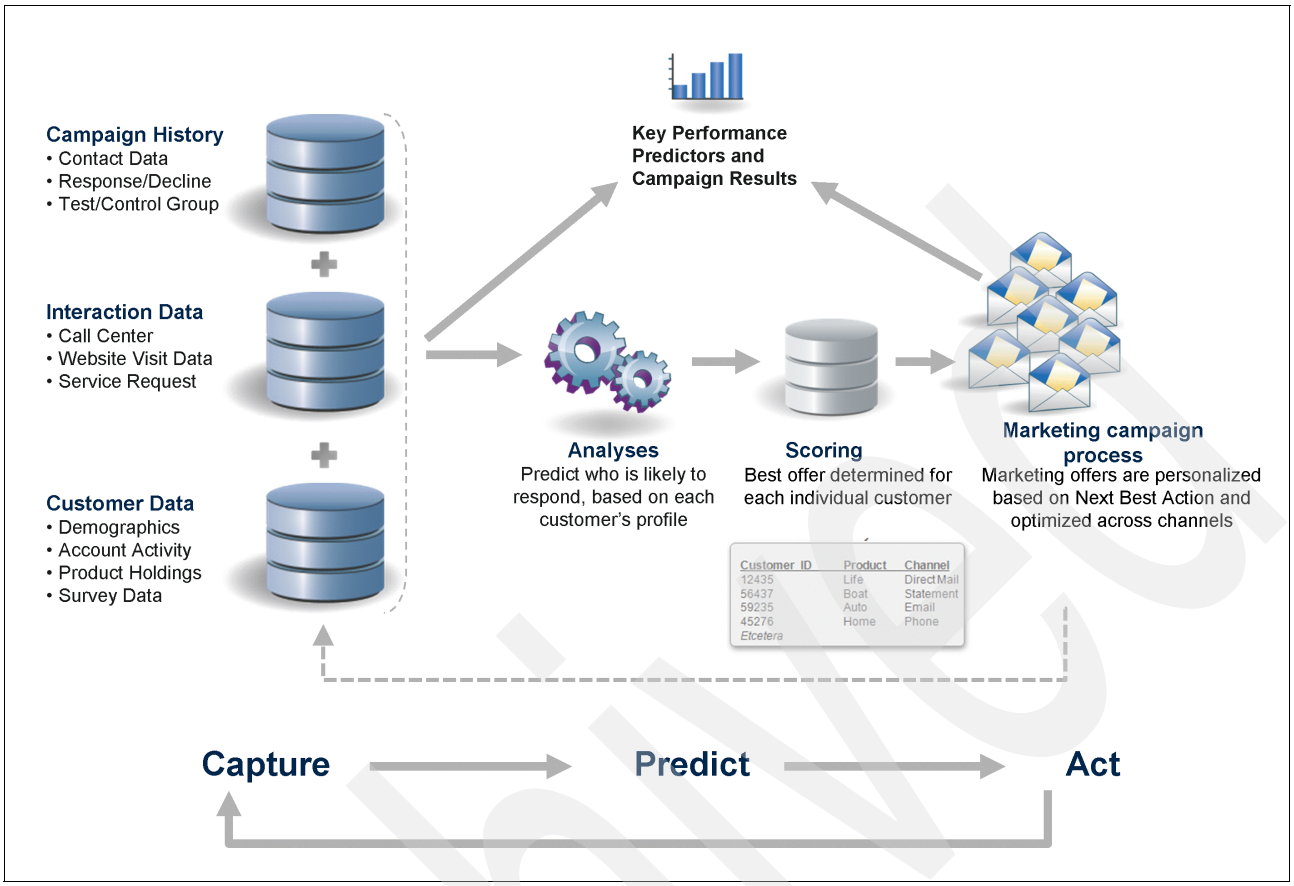

Figure 5 Analytical process, insurance customer retention and growth

IBM SPSS Predictive Analytics and Reporting for Claims solutions use advanced algorithms to analyze each claim to accelerate processing of valid claims, minimize fraud and unnecessary payments, and identify customized offers that create a more personalized customer experience.

In this solution, the first step is to collate multiple data sources such as customer data, interaction data, and campaign history, and internal expertise to create business rules for referral and fast-tracking. Claims and investigations data are analyzed to uncover additional rules, broadening the fundamental ability to identify high-risk claims. This data is input into predictive algorithms that, when combined with business rules, score claims and categorize them as low, medium, or high risk. Based on those scores, the appropriate action can be identified. In some cases, this can mean *fast-tracking* low-risk claims, which often results in increased customer satisfaction. In other cases, claims undergo further investigation and the customer is contacted through the appropriate channel for further claims processing.

Automating the assessment of each open claim using these rules and predictive models will, over time, improve detection rates and reduce false positive results. The reason is because each interaction (for example, claim processing) presents an opportunity to capture additional claims and investigation data, which strengthens the predictive models. The risk-based claims processing solution is depicted in Figure 6 on page 13.

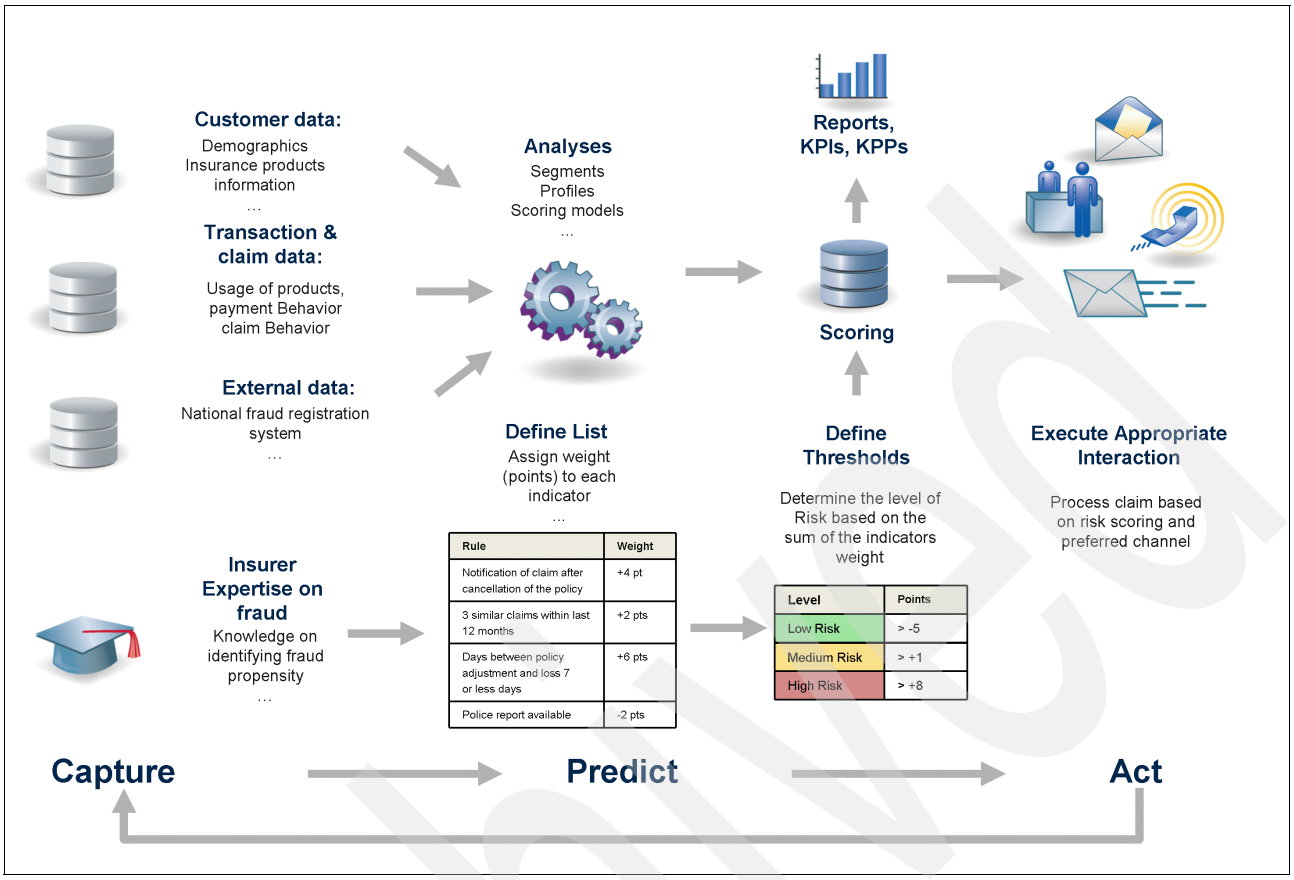

Figure 6 Analytical process, predictive analytics and reporting for claims

Telecommunications

This section describes the business and functional requirements of the telecommunications (telco) industry, and offers examples of how IBM SPSS Predictive Analytics can help those industry-specific needs.

Business and functional requirements

In the highly competitive mobile and cellular telecommunications market, where consumers have few barriers to switching carriers, churn is a major issue. Companies often detect the intent of customers to churn after it is too late to take action, making timeliness and the ability to preempt customer attrition a key goal for this industry. Additionally, the high cost of acquiring replacement customers, and the time taken for them to grow to the value of established customers, makes a retention strategy essential. Doing so involves identifying at-risk customers in tandem with identifying the appropriate remediation, and doing so at the individual, rather than aggregate, level.

Because of the dramatic cost of customer churn, fast deployment and quick time-to-value of a retention solution are critical goals of IT professionals and LOB managers.

Key focus areas: Churn reduction, marketing optimization for customer acquisition and up-sell

Predictive analytics approach

IBM SPSS Telco Churn Management solutions uncover key predictors that detect churn propensity well before the customers consider it, and then identify the correct action to take for each at-risk customer.

The solution focuses on the identification of at-risk customers at every point of interaction using the rich data that is already held by telecommunications companies about their customers. Examples of this data includes call detail records (CDRs), descriptive (demographic) data that is drawn from customer profiles at the time of contract initiation, and interaction data in the form of text notes from service-center conversations, which can also be key sources of attitudinal information regarding what motivates a customer to remain loyal.

The core model to be built is one that scores a customer on how likely that customer will cancel a contract within a given time from, such as 2 - 3 months, a period that allows sufficient time for retention actions to be taken. Targeting all at-risk customers, however, is expensive and can be counter-productive; certain mechanisms to measure customer value are critical in selecting only those customers most worth keeping. This approach might include a current measure of profitability, a traditional, static formula for calculating likely lifetime value (LTV), or a predictive model that estimates LTV dynamically for any customer, or some combination of these.

Action may be taken by scoring customers in sync with the monthly billing cycle and associated data refresh, on both their churn propensity and value. These scores and selections, along with model-selected treatments and under the control of rules governing marketing policies, are fed to campaign management systems that run outbound, proactive retention campaigns.

This solution becomes even more effective if sudden changes in defection risk levels of customers can be detected. To do this task, live models must be deployed to inbound channels such as the service center. The models then respond on a real-time basis, combining the latest behavioral data with additional information that is surfaced during the call and that can reveal imminent defection risk in what was previously thought to be a “safe” customer.

Because of the highly dynamic telecommunications market, models can obsolesce quickly if not updated with new data representing the latest customer behaviors and preferences. To manage this, models can be automatically monitored and refreshed so that they are brought up to date when their performance declines. The telco churn management solution is depicted in Figure 7 on page 15.

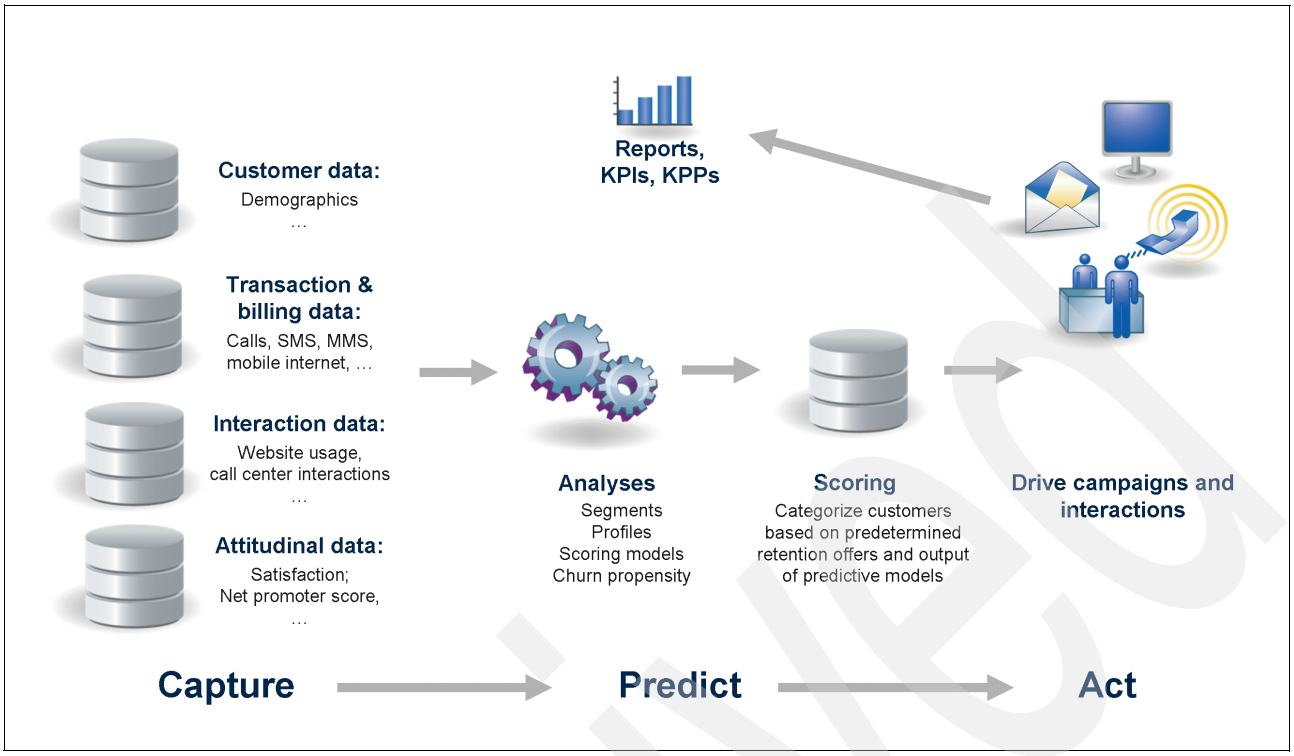

Figure 7 Analytical process, telco churn management

Retail

This section describes the business and functional requirements of the retail industry, and offers examples of how IBM SPSS Predictive Analytics can help those industry-specific needs.

Business and functional requirements

Retailers who run promotions on the wrong product combinations or whose offers to customers are poorly targeted, waste effort, money, and opportunity. Maximizing profitability in the retail segment requires deep understanding of customer preferences and product sales patterns to be able to identify effective product assortments. When selecting which products to offer, retailers are challenged with balancing inventory costs against lost sales opportunity, if desired products are not available. To maximize return on marketing investment (ROMI), retailers must design promotions that match shopper preferences and behavior and build customer loyalty through consistent service delivery, by offering the right products, and by promoting customer satisfaction.

IT professionals typically focus on using existing data more effectively, gaining better visibility into information with fewer IT resources. With the proliferation of online shopping, the ability to offer effective and appealing customer self-service capabilities is also critical. Retail marketers and customer loyalty executives seek to understand customer preferences and buying patterns, improve their ability to develop effective promotions, and boost marketing conversion rates per unit-spend.

Key focus areas: Marketing optimization for targeted promotions, customer loyalty, forecasting and assortment planning (optimized supply chain)

Predictive analytics approach

With *IBM SPSS Market Basket Analysis* solutions, advanced predictive algorithms automatically find product combinations that drive increased sales, and ensure that marketing offers are tailored to each customer's needs and preferences, increasing revenues and return on marketing spend.

Predictive models examine point-of-sale data from all customers to reveal which combinations of products tend to be purchased together. By examining these associations, retailers can determine which joint offers, of pairs or sets of products, are most likely to generate additional sales. This approach can provide certain in-store deployment capabilities, such as posting discounts when products are purchased together. Although this approach certainly provides several advantage, the effectiveness of it is entirely dependent upon customers entering the store and physically seeing the offers that are displayed.

When this information is combined with other customer data such as demographics, behavior, interactions and attitudes, purchasing patterns can be matched to customer profiles and enable the targeting of specific offers to specific customers. This second level of predictive models can then identify which items should be included in a promotional brochure to be sent to a customer, and by which channel it should be delivered (postal mail, email, and so on). Customer loyalty programs are often a key source of this type of individual customer data, as are customer satisfaction surveys or website activity. The retail market-basket analysis solution is depicted in Figure 8.

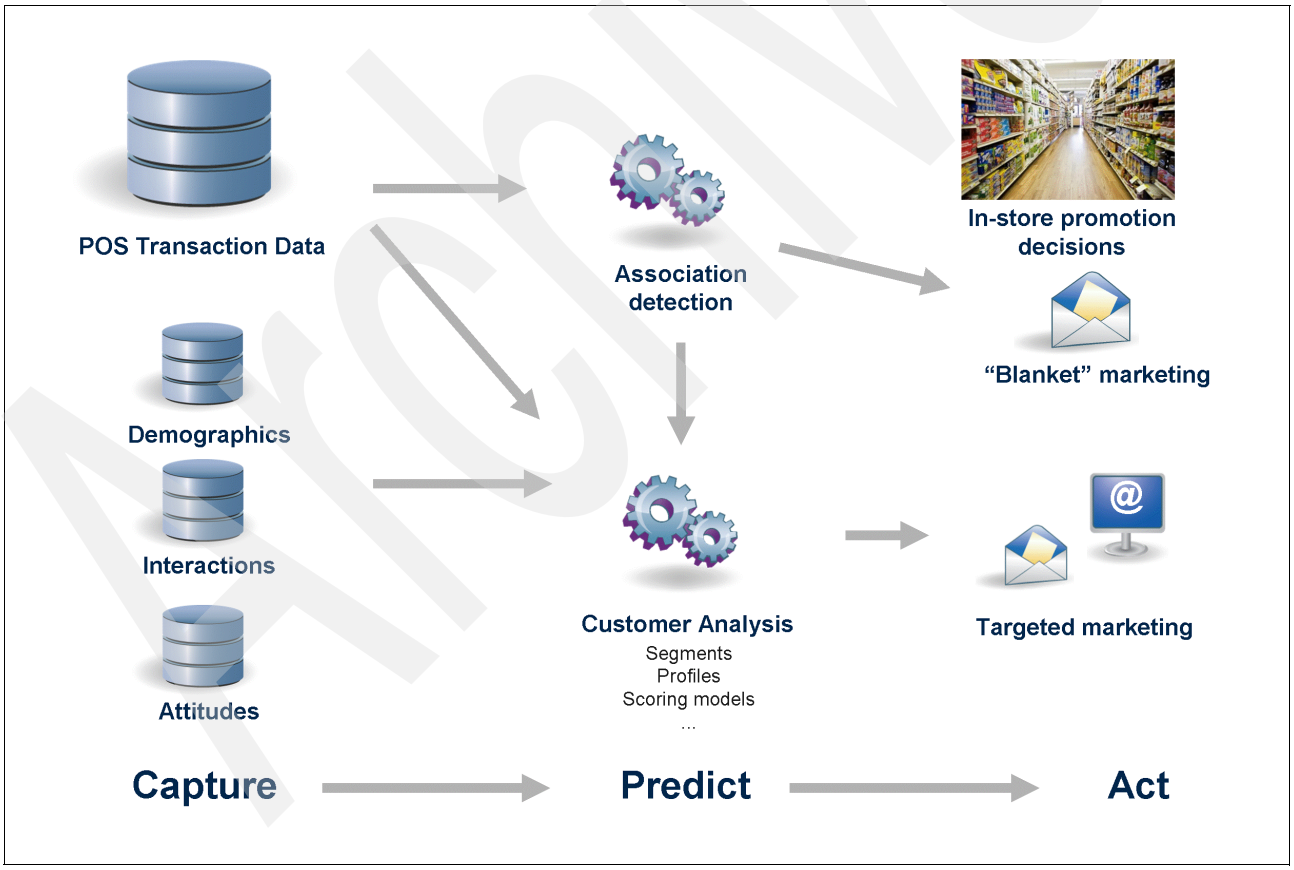

Figure 8 Analytical process, retail market-basket analysis

IBM SPSS Assortment Planning solutions predict the optimum stock levels and assortments at the store, category, or even SKU level (even in volatile environments) to maximize resources, increase inventory turnover, and increase customer satisfaction. This is about

predicting which items a customer will want to buy at a particular store, and ensuring the items are available. This approach includes techniques such as demand-based store clustering, new and existing SKU sales forecasts, category-space allocation, association modeling (market basket), and assortment optimization and rationalization.

Using predictive models, stores can determine how many assortments (store clusters) are needed for each product category. Clustered assortments allow for grouping of similar stores together to maximize customer traction and lower the demands on the supply chain. These clusters are demand-based, rather than demographic based, where the actual demand for key product SKUs is revealed, helping stores to more easily identify which products to stock.

The models also forecast unit sales for new and existing SKUs. Products are divided into attributes and assigned relative importance, and models identify key predictors for arriving at an accurate forecast. Predictive capability essentially makes it possible to look beyond historical sales as indicators of future ones. The retail assortment planning solution is depicted in Figure 9.

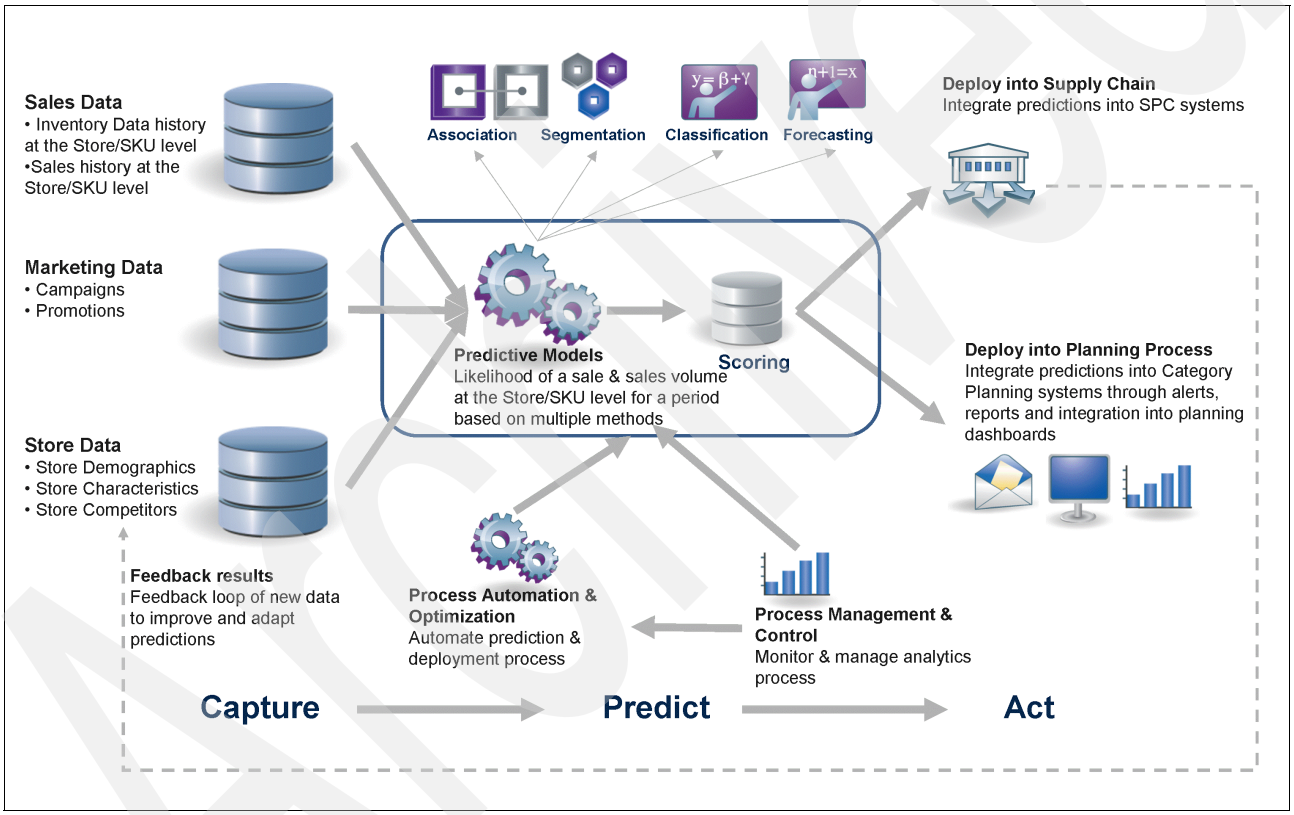

Figure 9 Analytical process, retail assortment planning

Government and public sector

This section describes the business and functional requirements of the public sector, and offers examples of how IBM SPSS Predictive Analytics can help address those industry-specific needs.

Business and functional requirements

The public sector includes such agencies as local crime prevention, federal border protection, healthcare, and education. Because these are not commercial enterprises, the organizational goals focus less on metrics such as explicit profitability and more on those such as crime

reduction, human resource management, operational planning, effectiveness of medical technologies, and effectiveness of educational resources (as reflected by student performance).

Technical challenges, however, remain similar: systems in place are antiquated and based on incompatible, siloed systems, preventing effective information-sharing across departments and a general lack of sufficient reporting and analysis capabilities. Many processes, such as pattern-determination across crimes, are often manual paper-based processes; within the healthcare realm, traditional reliance on opinions of well-trained health practitioners is favored over the use of objective data. The ability to effectively analyze data, understand events that trigger situations such as criminal activity, employee attrition, and even cardiac arrest does not exist. Real-time feedback and the ability to quickly disseminate and act upon information, is critical to dynamic, point-of-impact decision-making but is generally absent.

Key focus areas: Public safety and security, reduction of crime and fraud, human resource management and resource planning, effectiveness of education and other public services.

Predictive analytics approach

IBM SPSS Crime Prediction and Prevention solutions combine data from disparate sources and analyze crime patterns, enabling real-time decisions, such as preemptively deploying resources where crime is likely to occur. Unified reporting platforms also enable and encourage cross-organizational information sharing and more sophisticated crime intelligence.

Various predictors that indicate likelihood of crimes are uncovered by using IBM SPSS Modeler, which finds hidden relationships in data that might not otherwise be apparent. When the predictive model returns results that are positive for likely crime incidents, the officer on duty who is monitoring the real-time dashboard can radio the patrol officer to indicate the zones where the crime is likely to occur. Using a mapping tool on the computer in the squad car, the patrol officer then views the new patrol routes and alerts surrounding officers of the changes. The outcomes of these decisions are then captured, tracked, and incorporated into the models, which are then updated based in this new information, become even smarter.

Information from historical crime incidents, such as location, crime type, victims, suspects, and severity of the crime is mined from an existing records management system. *Enabling factors* such as weather can be obtained from the city's weather forecast provider. *Trigger events*, also information collected from the city, include holidays, events, festivals or paydays. Unstructured data can also be collected from crime reports, surveillance videos, and call recordings, which provide the critical counterpart to historical crime data and enrich the data set upon which the predictive models are based.

Beyond crime prevention, public safety and security organizations can also use business analytics to prevent both internal and external terrorist threats and decide where to locate emergency command centers based on the frequency and location of crimes, fires, and accidents. Predictive modeling helps determine which criminals are at risk of re-offense and need placement in more focused rehabilitation programs. In addition, agencies can benefit from using analytics for traffic risk profiling, suspect vehicle identification, and asset maintenance. The crime prevention and prediction solution is depicted in Figure 10 on page 19.

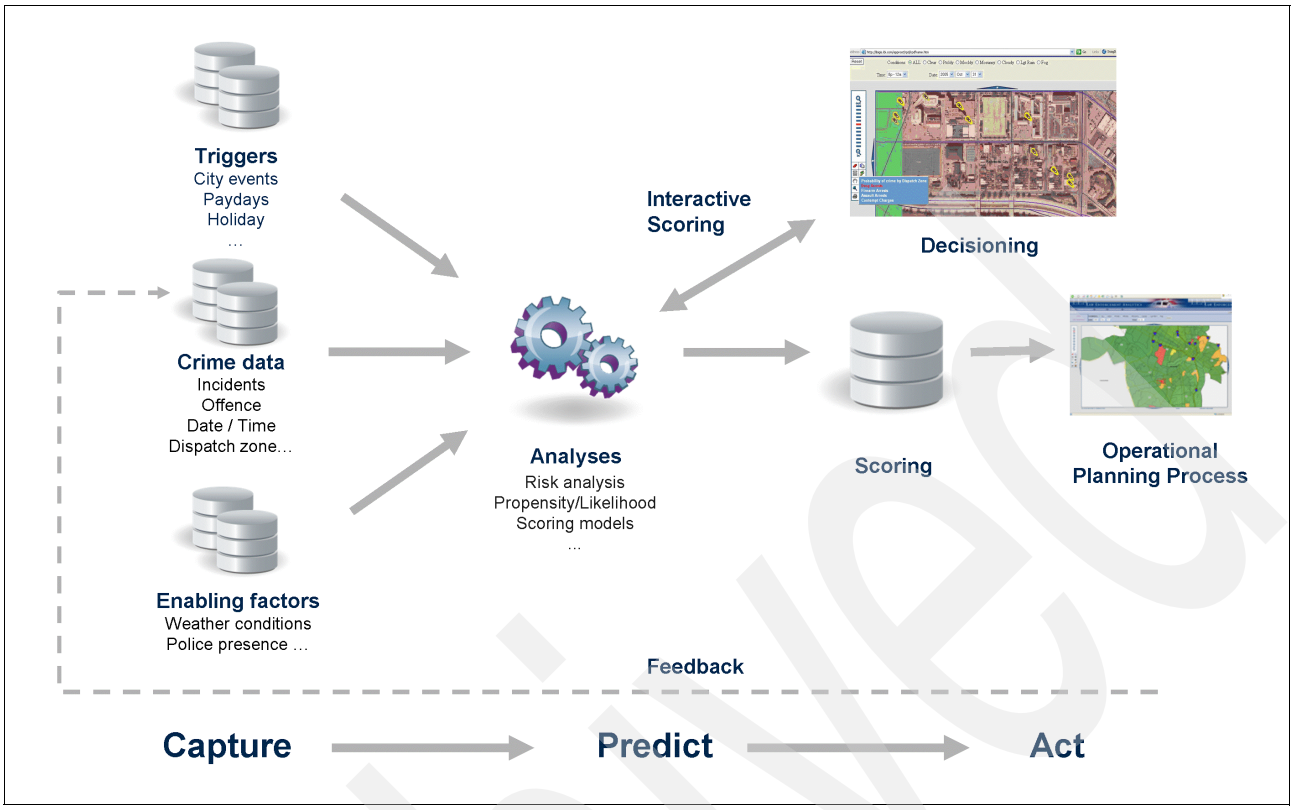

Figure 10 Analytical process, crime prevention and prediction

Within the education realm, predictive analytics can also help a school district's challenge with making static, point-in-time decisions about programming and failing to consider the dynamic needs of students, which results in wasted resources and sub-par student performance. *IBM SPSS Student Performance* solutions use advanced algorithms to adapt and match the right programs to each student throughout the year, improving standardized test scores, preempting dropouts, and creating a positive, customized experience for each student.

Industrial

This section describes the business and functional requirements of industrial organizations, and offers examples of how IBM SPSS Predictive Analytics can help address those industry-specific needs.

Business and functional requirements

Early identification of maintenance requirements and operational issues is critical to preventing production interruptions, improving usability and service levels for customers, and meeting and exceeding SLA expectations. It is the key to improving maintenance cycles and reducing costs.

The challenge for manufacturing has always been to produce high-quality goods while optimizing resources at every step of the process. Over the years, manufacturers have developed a number of sophisticated approaches to quality control, supply chain management and equipment maintenance. Today, with continuing cost-control efforts, plant managers, maintenance engineers, and quality control champions all want to know how to sustain quality standards and avoid expensive unscheduled downtime or equipment failure,

and how to control the costs of labor and inventory for maintenance, repair, and overhaul (MRO) operations. In addition, managers in finance and customer service and, ultimately, in the executive suite, have a stake in how well production processes deliver finished goods.

Key focus areas: Operational uptime, reduction in maintenance costs, supply chain management and optimization

Predictive analytics approach

IBM SPSS Predictive Maintenance solutions combine data from disparate sources and automatically detect failure patterns, enabling preemptive deployment of maintenance and repair resources and dramatically saving downstream costs.

A fully automated set of predictive models analyzes both structured and unstructured data in real time, and may include information about production assets, usage, environment, and previous maintenance. The models quickly detect failure patterns and identify the root cause of the problem, and also the likelihood that a similar event will happen again in the future. These models can be combined with domain expertise and optimization rules to improve maintenance and inventory planning.

Because engineers have 24x7 access to data on every piece of equipment, they can evaluate the reliability of every asset and build a maintenance schedule that performs inspections and maintenance in time to prevent failures. This technique eliminates the need for shutting down a line simply to perform regularly scheduled maintenance that might not really be necessary.

As operating conditions change, metrics for every piece of equipment are updated in real time. The advanced algorithms that are contained in the solution can determine the reliability of every asset at any time in the future, so that inspections and maintenance can be performed at exactly the right time. This predictive maintenance solution also identifies the replacement parts needed to support this more accurate maintenance schedule. It eliminates the need for unnecessary and expensive overstocking of spare parts. Manufacturers can now maximize both the allocated labor resources and spare part inventories. The result is the elimination of undue maintenance, prevention of costly down-time and repairs, and reduction in MRO inventory-carrying costs.

The predictive maintenance solution is depicted in Figure 11.

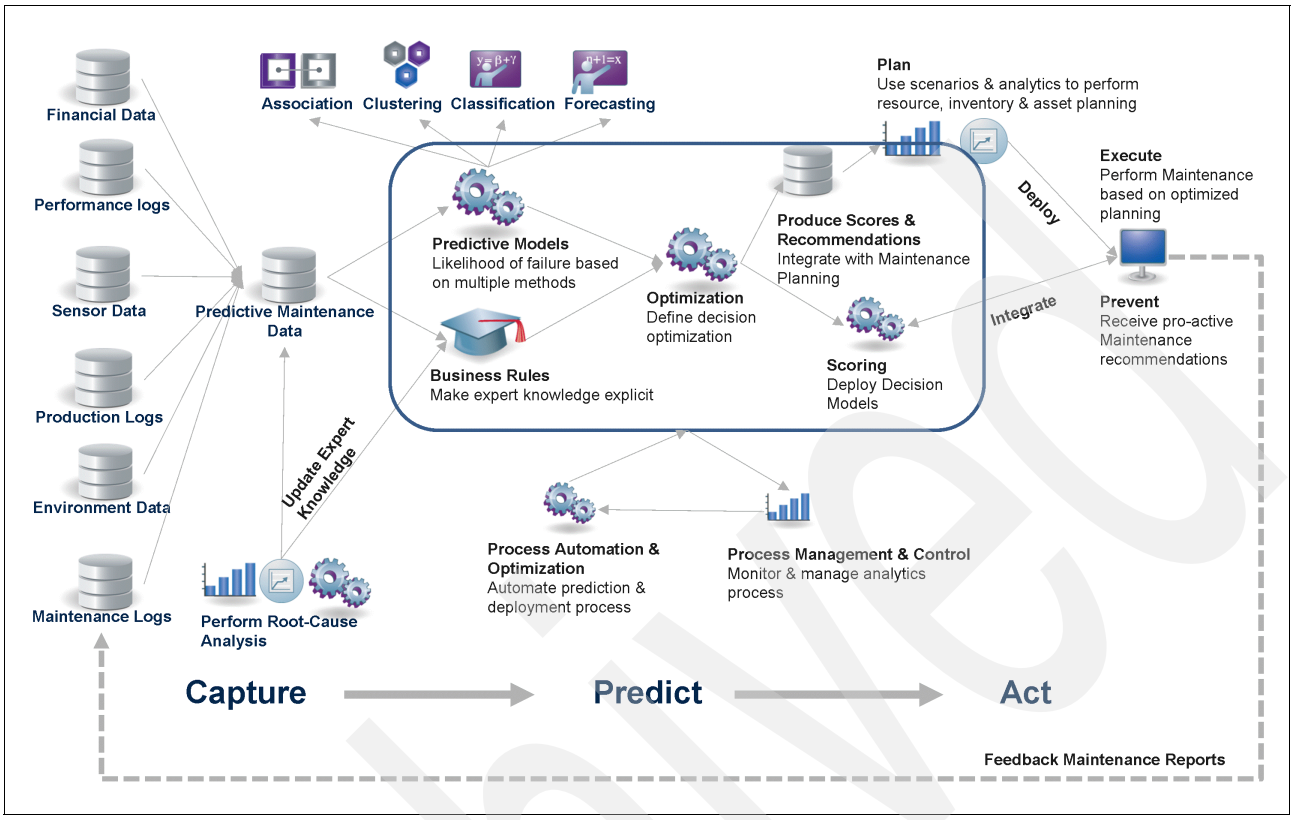

Figure 11 Analytical process, predictive maintenance

Components of a solution

To understand how predictive analytics delivers value, think about a predictive analytics solution as consisting of three main elements. This section describes those elements.

Data-driven business decisions

At the top level, predictive analytics enables better business decisions. Analysis might reveal new insights that can drive far-reaching, strategic decisions by senior management and deliver step changes in business value. However, a more common way is to see these insights applied at the level of individual cases, enhancing key business decisions that are made frequently and repeatedly, where improvement leads to a higher proportion of good outcomes and a clearly measurable, incremental ROI.

A helpful way to think of these decisions is *if we could make better decisions about <X>, we could deliver greater value by <Y>*. Several examples are as follows:

- If we could reliably predict which of our high-net-worth customers were likely to defect to a competitor, we could intervene and offer incentives to persuade them to stay loyal.
- If we knew how likely each of our customers would respond to a particular cross-sell offer, we could reduce the size (and cost) of campaigns by: not targeting people unlikely to respond; increase response rates and revenues; and avoid the situation in which customers who have no interest in the offer see it as junk mail.
- If we could accurately assess the risk of each insurance claim as we receive it, we could fast-track safe claims to reduce costs and increase customer satisfaction and loyalty, while increasing our fraud-detection rate by ensuring that our investigative resources focus on genuinely high-risk cases.

Deployment into processes and systems

In predictive analytics, *deployment* describes the integration of the results of analysis (for example, pre-calculated propensity scores or a “live” predictive model) with business processes and the operational systems that support them to make better decisions and drive deliver better outcomes.

In certain solutions, this approach might be simple: one point of integration in one process and system. One example is as follows:

- At a single point in the processing of tax returns, a predictive model scores every return on the likelihood of non-compliance, and adds those with high-scores to a list to be audited by an investigations team.

In other solutions, the approach might be more complex:

- In insurance risk management, models are integrated with the claims-handling process at the claim-creation step to assess risk and recommend a routing (fast-track, investigate, normal) for each claim. For claims following the normal process, the models can also be run at later stages in the process to re-assess risk (and reroute claims as appropriate) as new data becomes available. At all these stages, the model invocation is integrated with the claims handling system that supports the process.
- A model for up-selling might have a single, simple purpose, which is to recommend whether a specific product should be offered to a specific customer, but might be integrated with multiple processes and systems to ensure offers are made through the

most appropriate channels and touchpoints for each customer. In a bank, such a model can be used in several ways, such as those in the following list:

- In the direct marketing process, to select customers to receive outbound mailing or calls, integrated with a campaign management system
- Within the online banking system, making suggestions to customers who log on to service their account
- Coupled to the transaction handling processes, making offers when customers visit a branch or use an ATM

The analytical process

The foundation of the business solution is the *analytical process* that produces the results to be delivered to the point of impact. Although this process might be as complex as any given situation requires, the analytical process follows three basic steps:

- *Capture* a holistic data view of the customer (referred to as the *analytical data view*).

  This step normally consists of up to four types of data:

  - *Descriptive*: This type is usually a mix of self-declared information and externally-sourced geodemographics.
  - *Behavioral*: This type can be as simple as transaction records (who bought what and when), or details of how customers use a product or service. For a wireless phone operator, this type can include detailed information about calling patterns. For a credit card provider, the type can include details of how customers spend with their card and how they pay off their bills.
  - *Interaction*: This type is the details of how customers interact with the company through its various channels. This information might include details of website visits by registered customers (though mapped from click-level data to business-meaningful events that happen during visits), or unstructured text (emails to customer service contacts, or transcripts of call center conversations). Also in this category is data originating from customers interacting *with each other*. In the era of Web 2.0, this interaction includes discussions on forums and blog postings, and *social networks* representing how customers communicate with and influence others. Integrating such data into the analytical data view requires specialized technologies, such as text mining, web behavior analysis and social network analysis.
  - *Attitudinal*: This type refers to the needs, preferences, opinions, and desires of people. In the past, these items have often been collected in the course of surveys that were conducted for market research or to assess customer satisfaction. However, in these cases the data that is collected has been used only in an anonymized, aggregated form. Today, many organizations recognize the value of using this data at the individual customer level, linked to the other elements of the analytical data view.

- *Predict* outcomes in current or future cases by analyzing the data, using advanced algorithms, to create predictive models.

  A prediction is the *raw* output of a model. It can be the propensity for a customer to behave in a particular way (for example, the likelihood that a customer who owes money will pay it back), or a selection of one of several options (for example, which of the available collection strategies will be most effective in persuading the debtor to pay).

- *Act* on the results of analysis. This involves the following actions:
  - *Decision management*: combining the results of possibly multiple models with business logic (rules, policies, exclusions etc.) to arbitrate between different possible actions and decide on the right one to be taken.
  - Integrating these decisions and actions into key points in the relevant business processes and the operational systems that support them.

Capture, predict, act takes us from collected data, through advanced analysis, to the successful deployment of analytical results to improve business processes. Note however, that this cycle is actually a *virtuous cycle*, producing continuous improvement: new data captured at the point of action (that is, during customer interactions) enhances the analytical data view, enabling more accurate predictions and driving better decisions with a greater proportion of positive outcomes and, hence, higher returns.

In many cases, elements of the analytical process can be automated to enable this iterative approach and to address other requirements such as monitoring of model performance and scalability to multiple related application areas.

IBM SPSS Software suite

The IBM SPSS product portfolio is represented in Figure 12.

Figure 12 The IBM SPSS product portfolio

The portfolio is designed to serve the three main phases of the analytical process, *capture, predict,* and *act*, and is structured into four main product families:

- Data Collection

  Delivers an accurate view of customer attitudes and opinions by collecting data from all customers and constituents, including that in multiple formats and languages.

- Statistics

  Drives confidence in results and decisions by verifying hypotheses and providing data-based evidence.

- Modeling

  Brings repeatability to ongoing decision-making with the ability to make accurate predictions.

- Deployment

  Maximizes the impact of analytics in your operation.

These categories are further described in the next sections.

Data Collection

This section provides a product overview, product highlights, and supporting images of the IBM SPSS Data Collection product family.

Product overview

IBM SPSS Data Collection is a complete suite of products for survey, market, or business researchers. It enables you to quickly and efficiently acquire clean data from the widest range of sources by using an expansive array of methods, and actively bring data about people's attitudes and preferences into your analytical decision-making. It is the best way to capture a complete perspective about your important constituents, making research efforts more accurate and more efficient.

This is increasingly important as business today demands faster, more representative and more cost-effective surveys for deeper insight into thoughts and opinions of customers. Both commercial organizations and market research firms rely on data collection's advanced technologies.

The IBM SPSS Data Collection family consists of the following components:

- Authoring

  Authoring helps you to streamline the creation of surveys, using familiar and intuitive interfaces. It also helps you to incorporate sophisticated routing and logic to increase completion rates and ensure clean, high-quality data for analysis and reporting.

- Interviewing

  Interviewing uses advanced technology for deploying and managing compelling surveys by phone, the web, or in person. It helps you to stay current with emerging trends such as SMS messaging or interactive voice response (IVR).

- Reporting

  Reporting helps you efficiently develop professional, interactive reports in online or desktop environments. It can convert insights into action by delivering the right information to the right people at the right time.

- Management

  Management helps you centralize the management of surveys and data with tools for all stages of the research life cycle, and increase efficiency and staff utilization by automating processes. As a result, tasks are done in the background and are ready when needed.

Product highlights

With IBM SPSS software, researchers can quickly and efficiently acquire clean data from the widest range of sources by using an array of methods.:

- A versatile platform integrates the full range of modern surveying methodologies so that researchers can create, field, and collect data simultaneously by using any mode, many languages, and from any location.
- The technology supports virtual call centers through Voice over Internet Protocol, which can be used to connect remote workers and within call centers.
- With IBM SPSS text analysis capabilities, researchers can quantify text responses for analysis with other survey data.
- Multilingual capabilities enable questions and instructions in almost any language.
- By fusing research information and client enterprise data, agencies can use predictive analytics to provide a more strategic service.

Sample images

This section shows several sample images of IBM SPSS Data Collection software, including the Author Interface, Data Collection Developer Library, Defining Project Templates,

Interviewer Administrator Screen, and various statistical outputs that are used during the data collection process.

The SPSS Data Collection main menu (Figure 13) shows the main activities: Design, Manage, Data Collection, and Reports. From here, you control all administrator functions such as launching a survey and controlling who may access which parts of the survey.

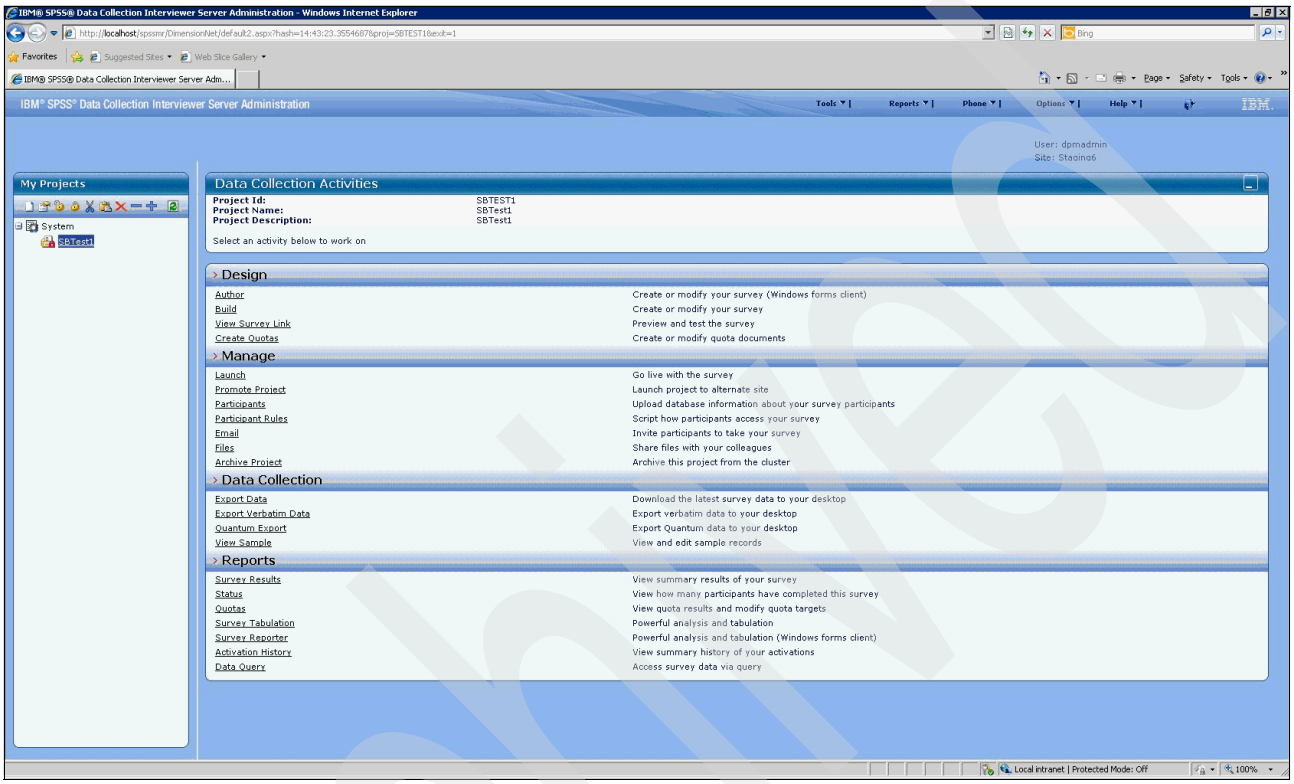

Figure 13 SPSS Data Collection main menu

The SPSS Data Collection Author interface (Figure 14) is used to create surveys.

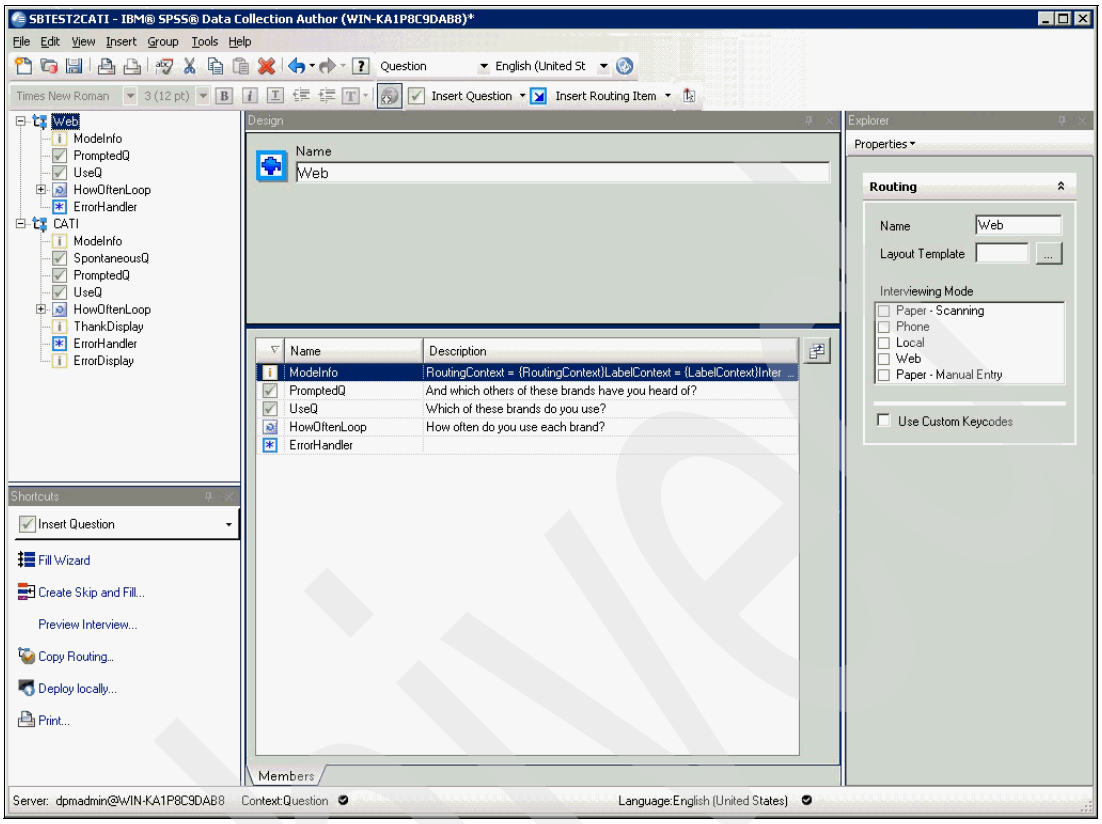

Figure 14 SPSS Data Collection Author interface

The SPSS Data Collection Computer Assisted Telephone Interviewer (CATI) provides a script to guide the survey (Figure 15).

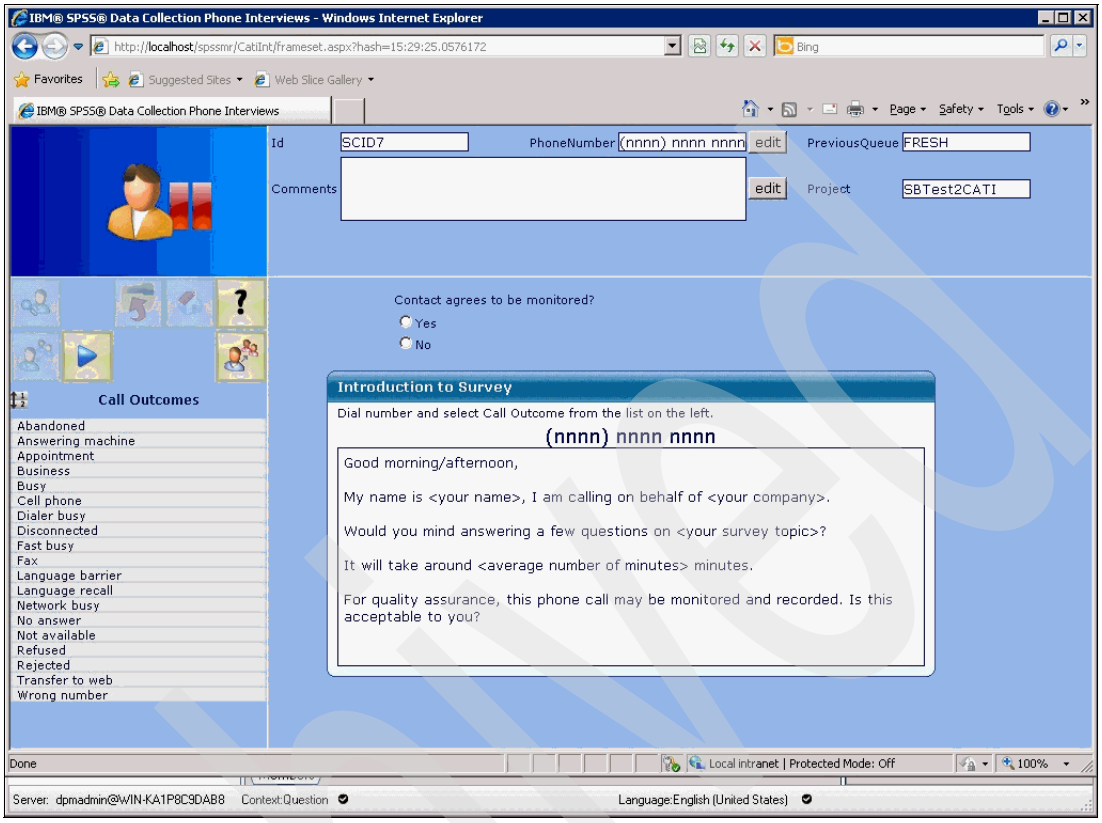

Figure 15 SPSS Data Collection Computer Assisted Telephone Interviewer

The SPSS Data Collection Computer Assisted Telephone Interviewer (CATI) prompts each survey question (Figure 16).

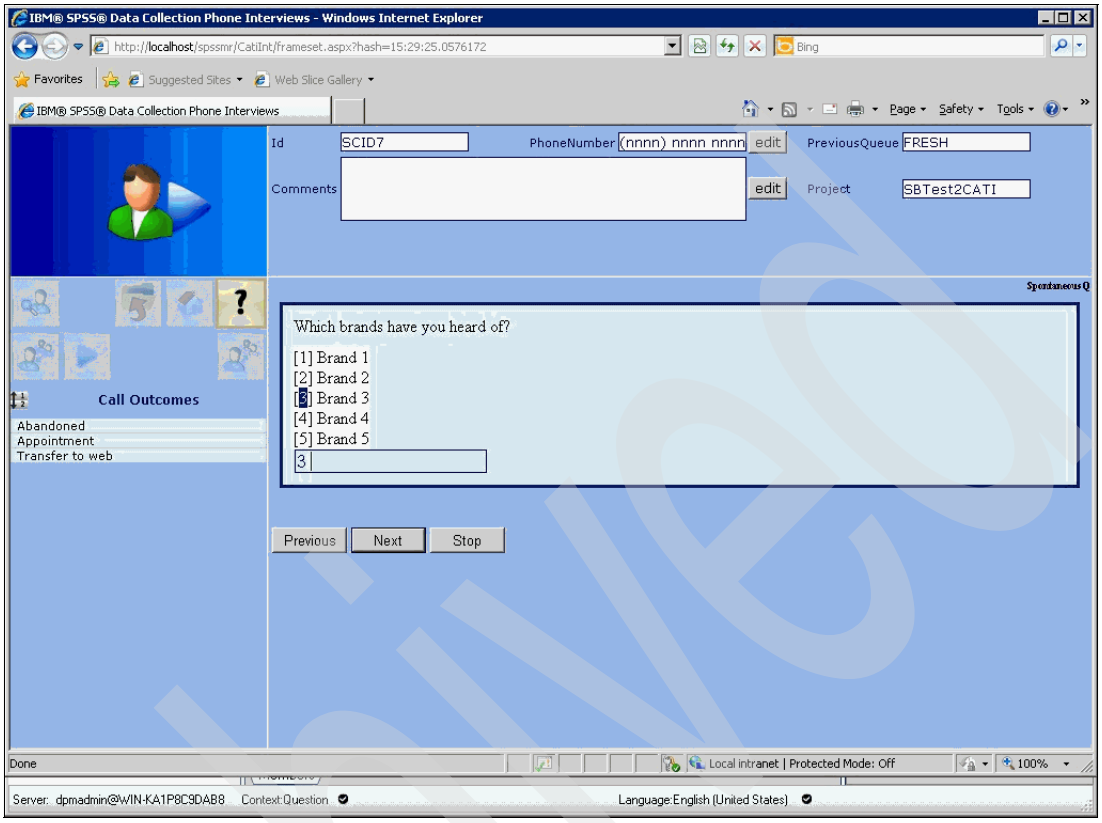

Figure 16 SPSS Data Collection Computer Assisted Telephone Interviewer sample questions

Defining Project Templates is new in SPSS Data Collection 6.0 and allows you to maintain standards across various projects (Figure 17).

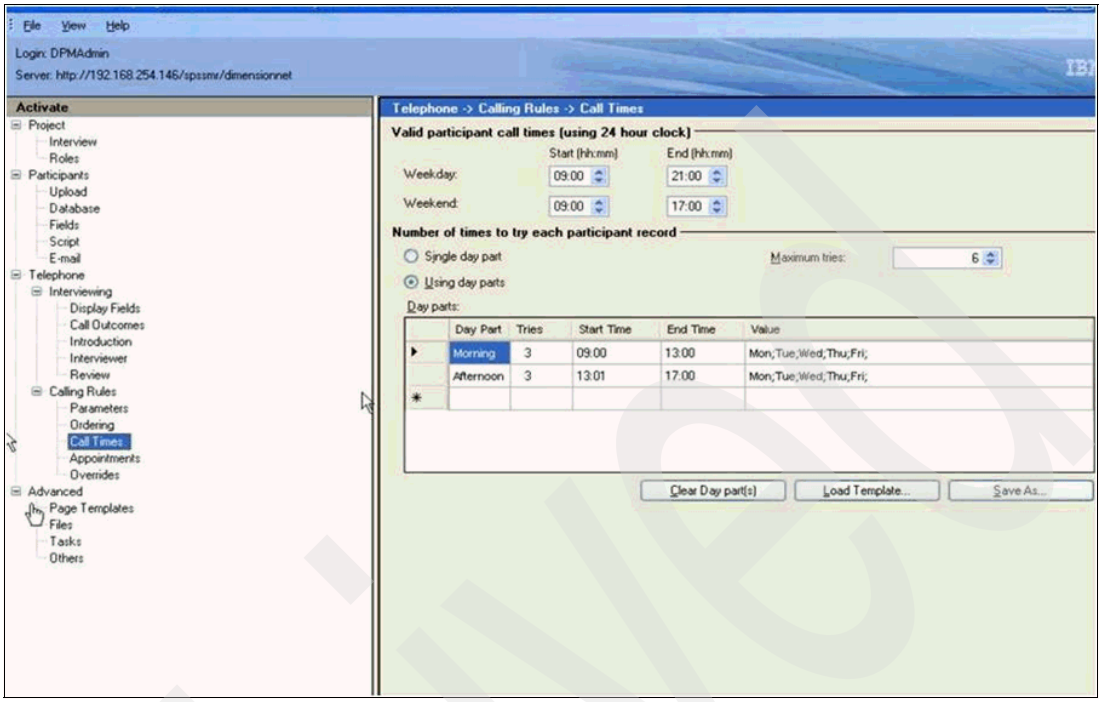

Figure 17 Defining Project Templates

The Custom Tables procedure enables you to create complex tables, such as the one shown in Figure 18. This table illustrates the relationship between the customers in each cluster and overall satisfaction, and includes several nested questions.

			TwoStep Cluster					
			Frequent BSI Road Warriors		Less Frequent Travelers		Loyal BSI Road Warriors	
			Would you stay with BSI again?		Would you stay with BSI again?		Would you stay with BSI again?	
			Yes	No	Yes	No	Yes	No
I am a satisfied BSI customer	Use spa facilities?	No	3.9	2.0	3.8	2.0	4.0	2.5
		Yes	3.6	.	3.6	2.0	3.6	1.0
	Ever called BSI to resolve a problem?	Yes	3.8	2.0	3.9	2.0	3.6	2.0
		No	3.6	2.0	3.6	2.0	4.2	1.7

Figure 18 Custom tables

Statistics

This section provides a product overview, product highlights, and supporting images of the IBM SPSS Statistics product family.

Product overview

IBM SPSS foundational technology uses sophisticated mathematics to help researchers validate assumptions and test hypotheses. From testing opinions about the latest product feature ideas or the viability of a political candidate, to the efficacy of a new drug treatment or prospective supply-chain allocation, statistics enables an organization to look at the beliefs of an organization and validate whether those views are based in fact. Gut feeling and instinct

are only as good as the experiences brought to the table; statistics can give you confidence in the results and final outcomes of decisions you make.

IBM SPSS Statistics is a leading statistical software suite that is used by commercial, government, and academic organizations to solve business and research problems. IBM SPSS Statistics is one of the most accessible statistics tool in the market, enabling organizations to apply mathematical discipline to their decision-making. Figure 19 shows capabilities.

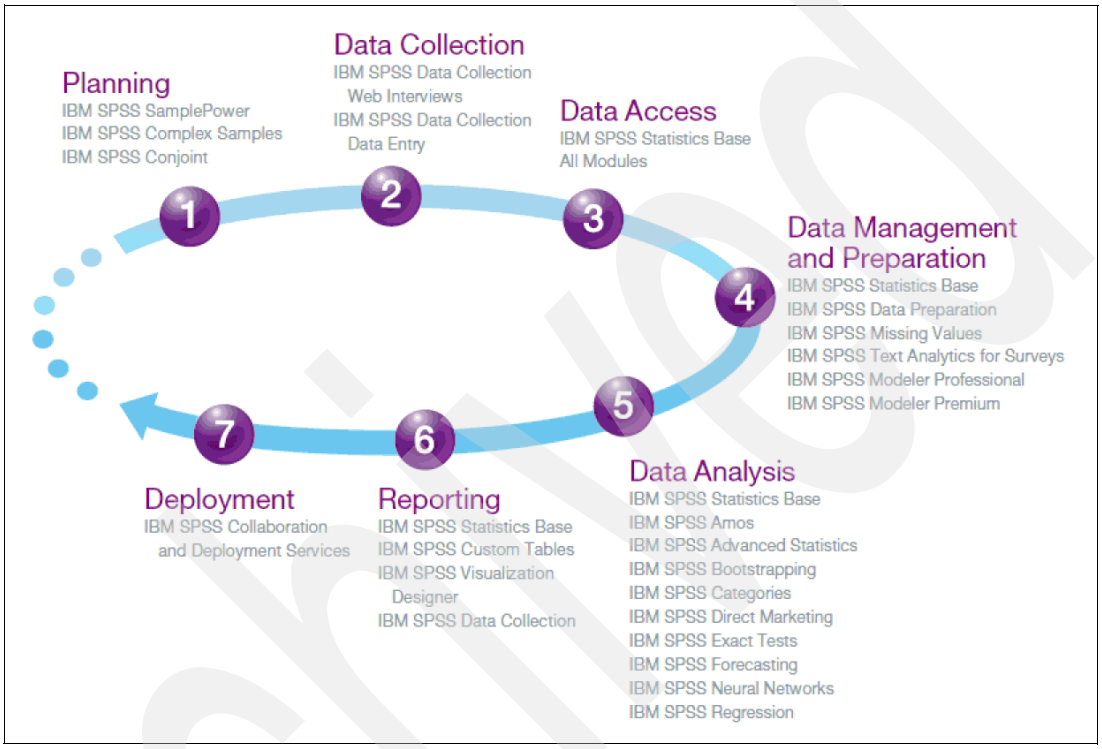

Figure 19 SPSS Statistics capabilities

Product highlights

Key benefits of using IBM SPSS Statistics are as follows:

- Multiple interfaces to key functions provide flexible access to a depth of analytical power.
- Comprehensive set of statistical tools enables the right statistical capability at the right time.
- Multiple deployment capabilities make the most of analytical resources.

Sample image

Figure 20 shows a menu of the analysis options in the SPSS Statistics software.

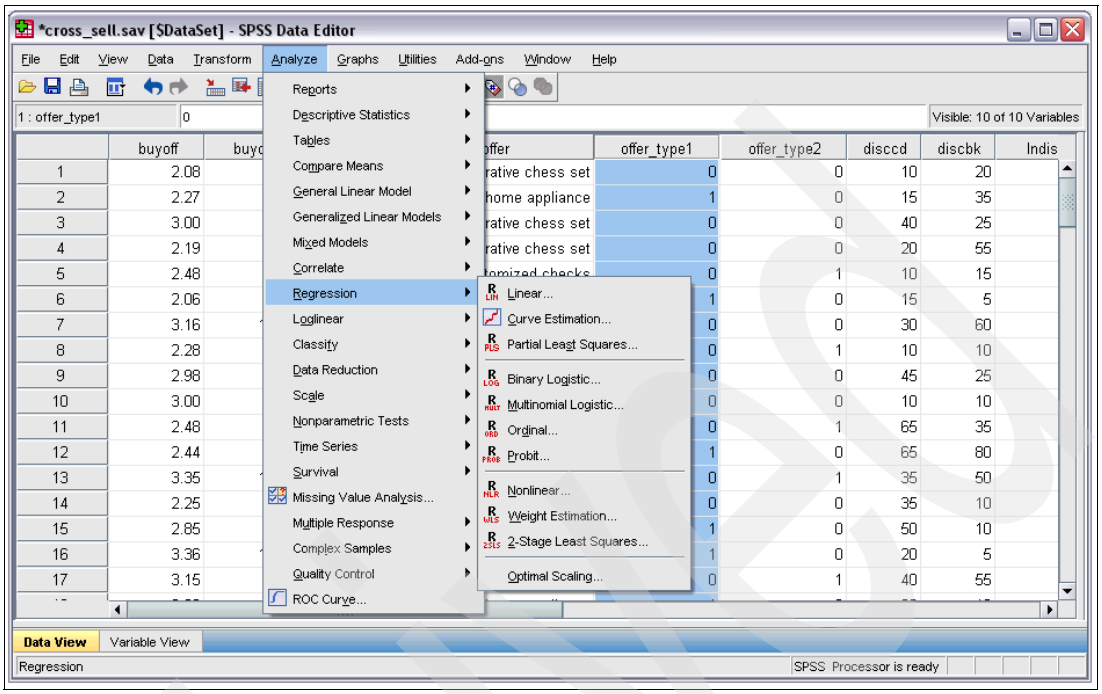

Figure 20 SPSS Statistics analysis options

Modeling

This section provides a product overview, product highlights, and supporting images of the IBM SPSS Modeling product family.

Product overview

While the powerful predictive analytics of Statistics is applied to testing hypotheses, the modeling family is the set of technologies designed primarily for creating models that predict future outcomes. Modeling, also known as data mining, helps organizations take seemingly unrelated data and find hidden relationships in data. Using these *models*, an organization can *look into the future* and understand what will happen in any current or future case based on what has happened before. From predicting which offer will have the most impact, to understanding and preventing churn, the modeling family helps people consistently make decisions, maximizing the results. This process repeatability makes modeling a powerful tool for embedding best practices inside the systems and processes of a business.

IBM SPSS Modeler is a powerful, versatile data and text analytics workbench that helps analysts build accurate predictive models quickly and intuitively, without programming. By using leading data mining techniques, the workbench helps analysts model future customer behavior, making predictions about how customers will behave. It is the most efficient way for organizations to use their data to model future customer behavior, predicting what that customer will do next. This technique is done by discovering patterns and trends in structured or unstructured data more easily, using a visual interface that is supported by advanced analytics. In addition to predicting outcomes, models can explain what factors influence them so users can take advantage of opportunities and mitigate risks.

IBM SPSS also offers integrated text analysis capabilities that extend the power of its key predictive tool and use the critical insights locked in unstructured data in a robust manner.

Text analysis increases the accuracy of predictions about customer behavior and strengthens the modeling capability.

Product highlights

Key benefits of using IBM SPSS Modeler are as follows:

- Access, prepare, and integrate structured data and text, and web and survey data.
- Support the entire data mining process with a broad set of tools that are based on Cross Industry Standard Process for Data Mining (CRISP-DM) methodology.
- Identify and extract sentiments from text in more than 30 languages and use this insight to build more accurate predictive models.
- Deploy textual insights so your entire organization benefits from a comprehensive, 360-degree view of the people you serve.

Sample images

This section shows several sample images of the IBM SPSS Modeler software, including examples of interactive visualization, automatic data preparation, and ensemble modeling.

IBM SPSS Modeler includes advanced, interactive visualization for models that use a single technique, or ensemble models that combine techniques making modeling results easy to understand and communicate (Figure 21).

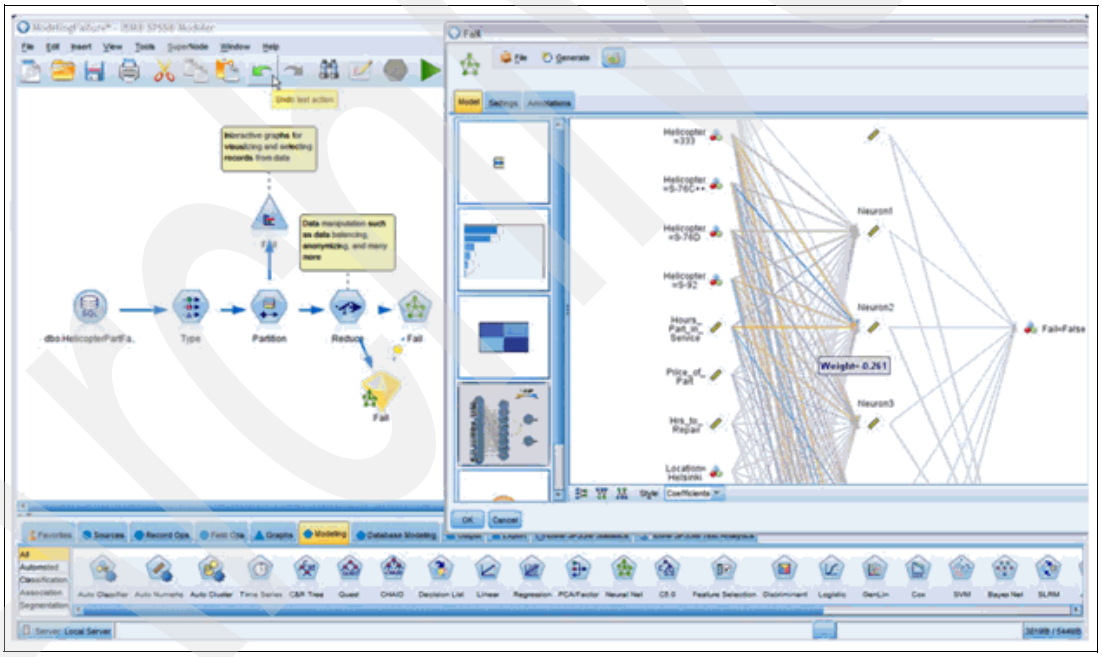

Figure 21 SPSS Modeler interactive visualization

IBM SPSS Modeler delivers leading ease-of-use features such as automatic data preparation and automatic modeling, helping you to build models that use single or multiple (ensemble) techniques. See Figure 22.

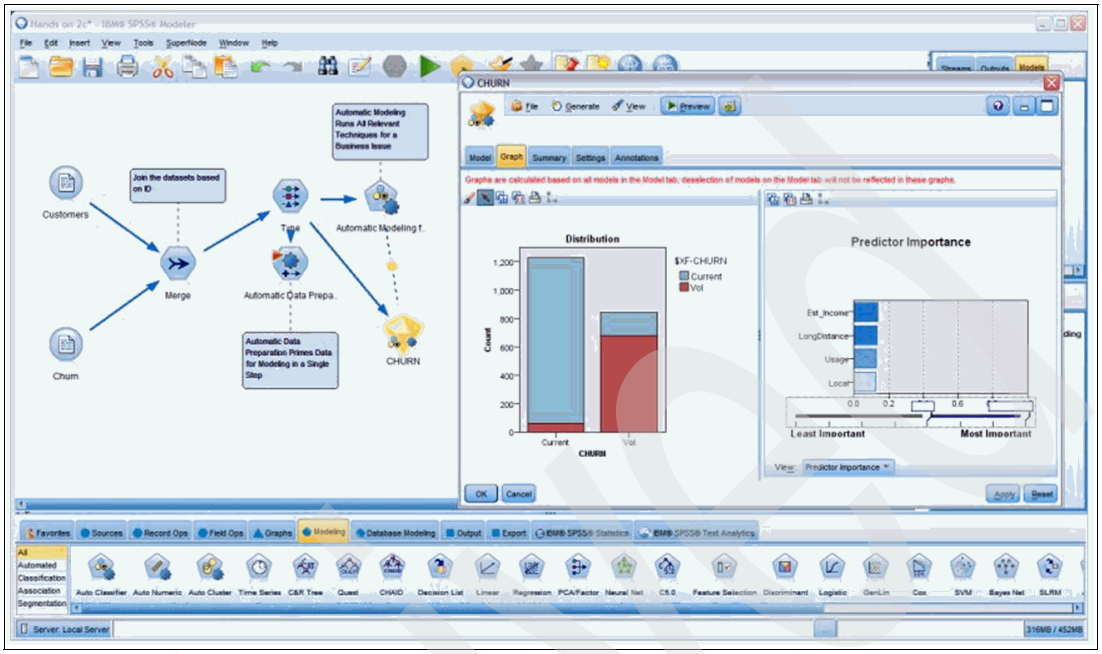

Figure 22 SPSS Modeler data preparation and automatic modeling

Deployment

The deployment family can help you maximize the impact of analytics in your operation by embedding the results of your analytic efforts in the hearts of enterprise systems. Deployment is about making analytics practical for the people who handle real-world challenges every day. From helping the call center agent by alerting them to the risk of a churning customer, to recommending corrective action for a failing student, the deployment family ensures that the processes of your organization operate at peak efficiency and that objectives are met.

The deployment family includes the following products:

- Decision Management
- Collaboration and Deployment Services

Decision Management product overview

IBM SPSS Decision Management combines predictive analytics and business rules to optimize and automate the millions of small decisions made every day. Users increase the positive outcomes of good decisions, and minimize the negative impact of bad decisions.

SPSS Decision Management harnesses the power of a variety of technologies that IBM offers, such as data mining, business intelligence, rules, event processing, data management, and then blends them all together. It is a great *mashup*. It facilitates the ability to create web-based, business user applications that are designed for a specific business problem. These applications help business people participate in the use of predictive analytics to meet their challenges.

No longer is the business solely reliant on back-office data analysts, data miners, or expert statisticians. The business can now use the power of predictive analytics and other critical enterprise technologies to create models to optimize outcomes at the point of interaction. All

this technology is wrapped in a graphical user interface (GUI), using language that is familiar and meaningful to the business user. Business can automate and optimize everyday decisions through a completely configurable solution template so that every decision is tailored and focused on the point of interaction (for example, website or contact center).

A key advantage is that the insight and skills of the analyst are not lost, because after the initial model is built by the business user, the analysts can apply their technical skills to fine-tune the results. The best part is that the insight gained from these decisions can be rolled back into the enterprise, creating a virtuous cycle or a closed loop system in which decisions grow more effective with each iteration.

The SPSS Decision Management architecture, which is part of the Collaboration and Deployment Services, is depicted in Figure 23.

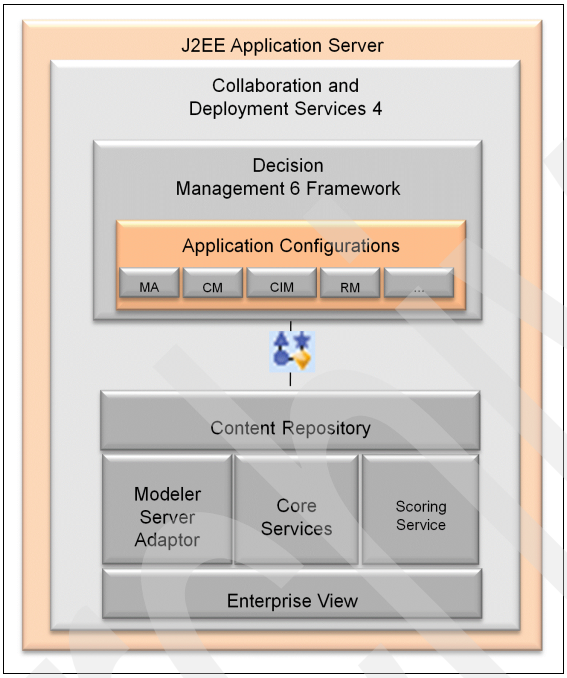

Figure 23 SPSS Decision Management architecture

Decision Management highlights

Decision Management is a highly effective method for optimizing and automating your business decisions. Supported by predictive analytics, it offers organizations the ability to move beyond reactive decisions to anticipate which actions are most likely to create successful outcomes in the future. Using Decision Management offers the following benefits:

- Include business knowledge in the modeling process.
- Create accurate models in a few simple steps.
- Get accurate results, faster.
- Improve collaboration, reuse and decision-making.
- Achieve the best possible outcomes for your business.

Sample images

This section shows several sample images of IBM SPSS Decision Management product, including the architecture and five steps comprising the Customer Interaction Management Module.

Figure 24 shows the interface for selecting data within Decision Management.

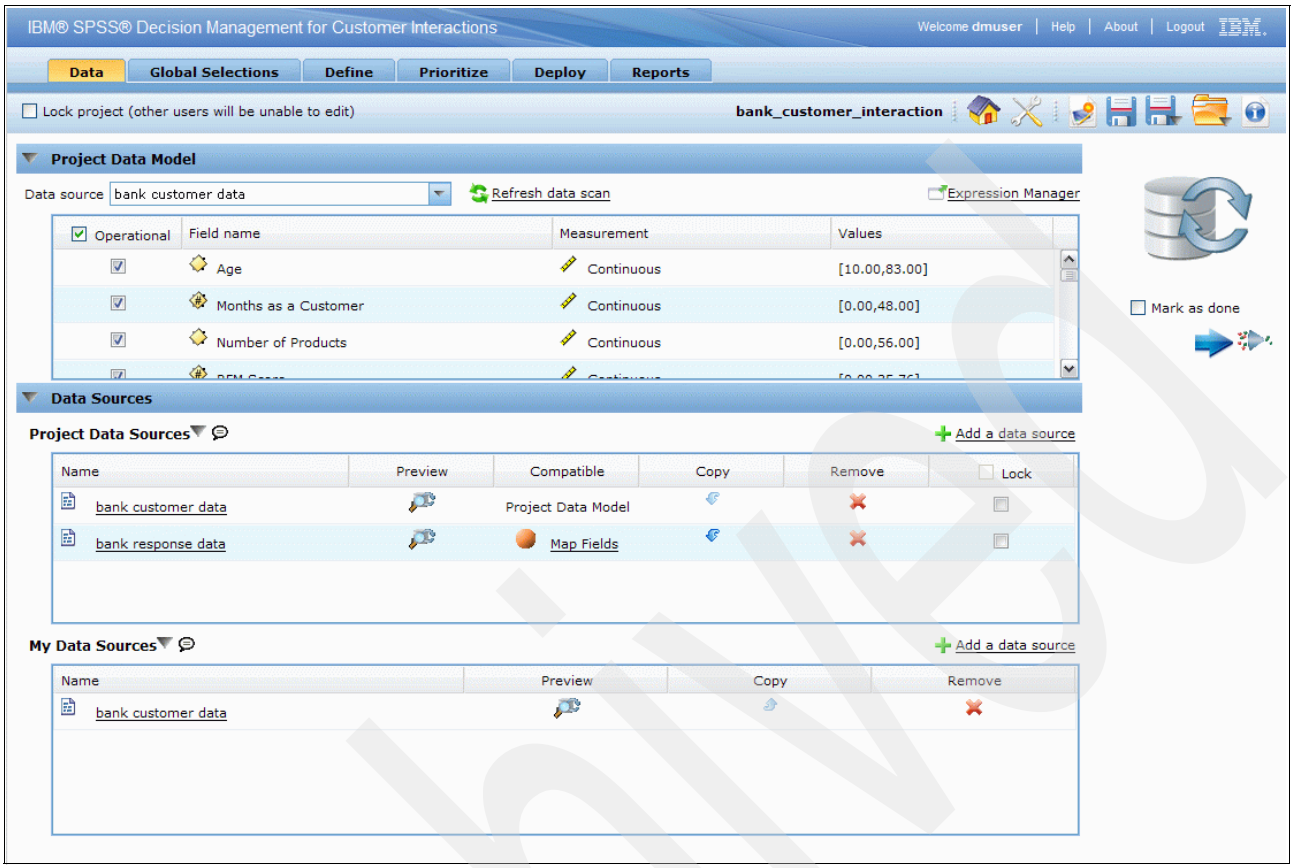

Figure 24 Selecting data within SPSS Decision Management

SPSS Decision Management allows organizations to make more targeted decisions that support business policies, such as which customers should receive marketing offers. Figure 25 shows an example of identifying bad payment customers and those who have recently received a separate marketing offer, and excluding them from upcoming campaigns.

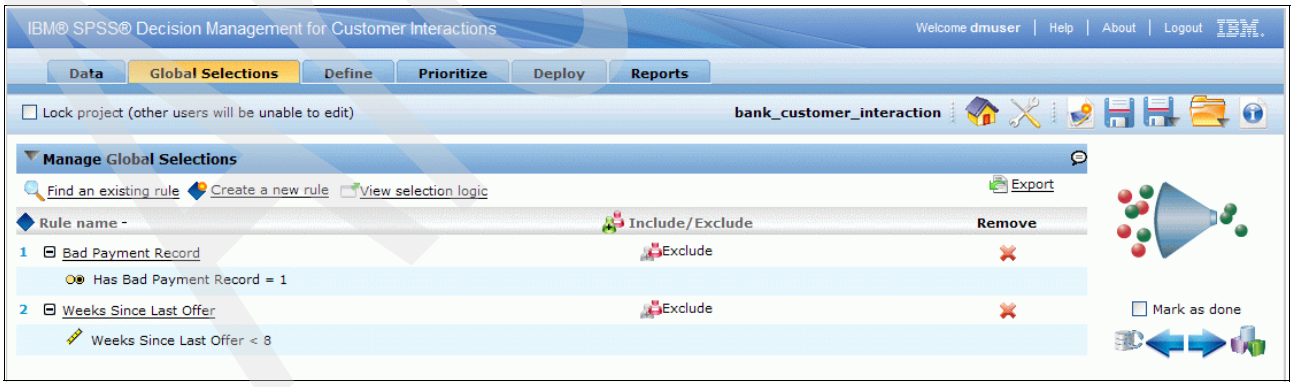

Figure 25 Filtering data within SPSS Decision Management

SPSS Decision Management combines existing business rules with the power of predictive modeling, as shown in Figure 26.

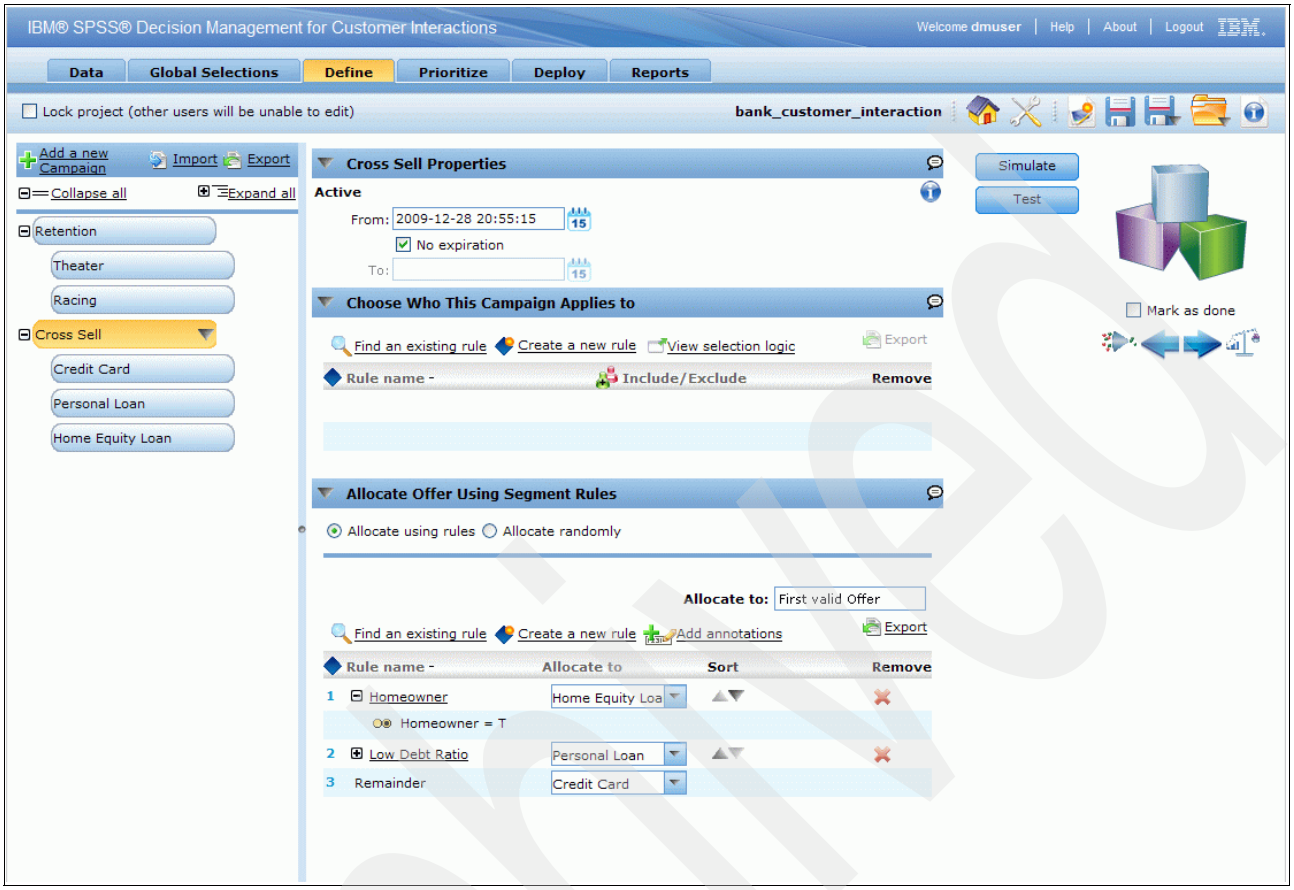

Figure 26 Creating rules and models within SPSS Decision Management

Based on the customer profile, likelihood to respond, maximum profit and other factors, SPSS Decision Management chooses the most appropriate response for that customer, as shown in Figure 27.

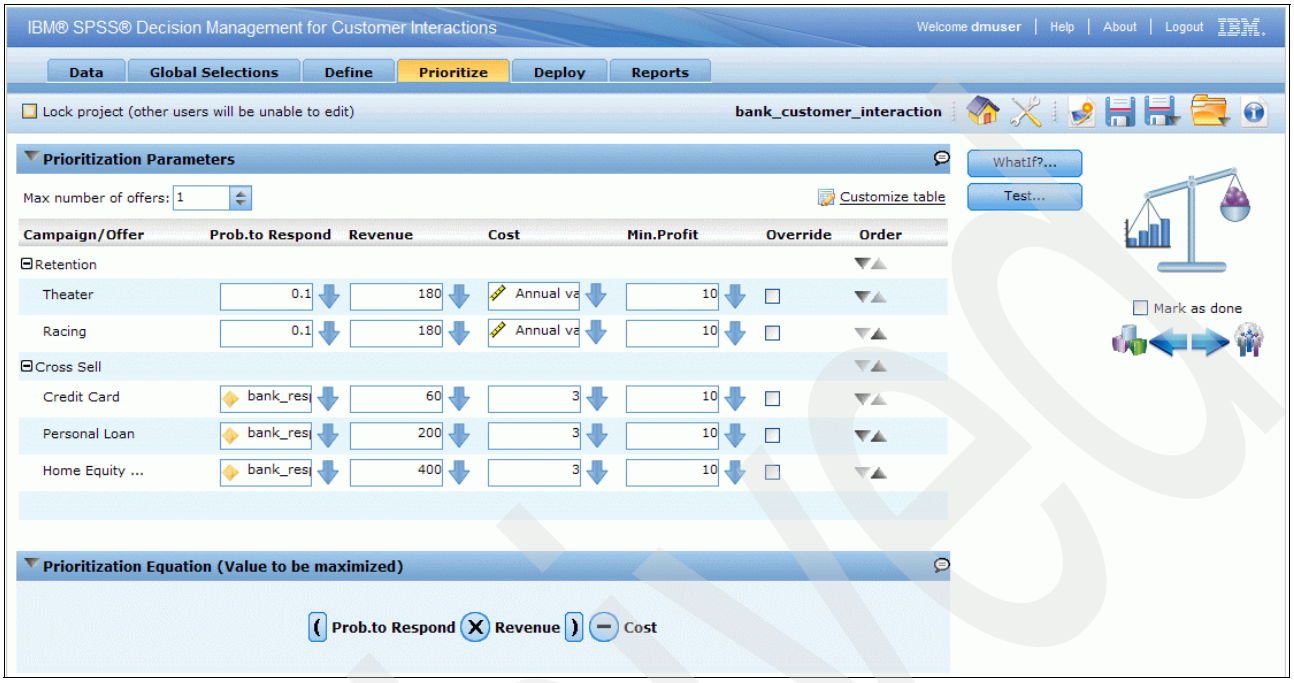

Figure 27 Prioritizing responses within SPSS Decision Management

Before deploying to production, customized questions can be created for users to ensure consistency and accuracy of data, as shown in Figure 28.

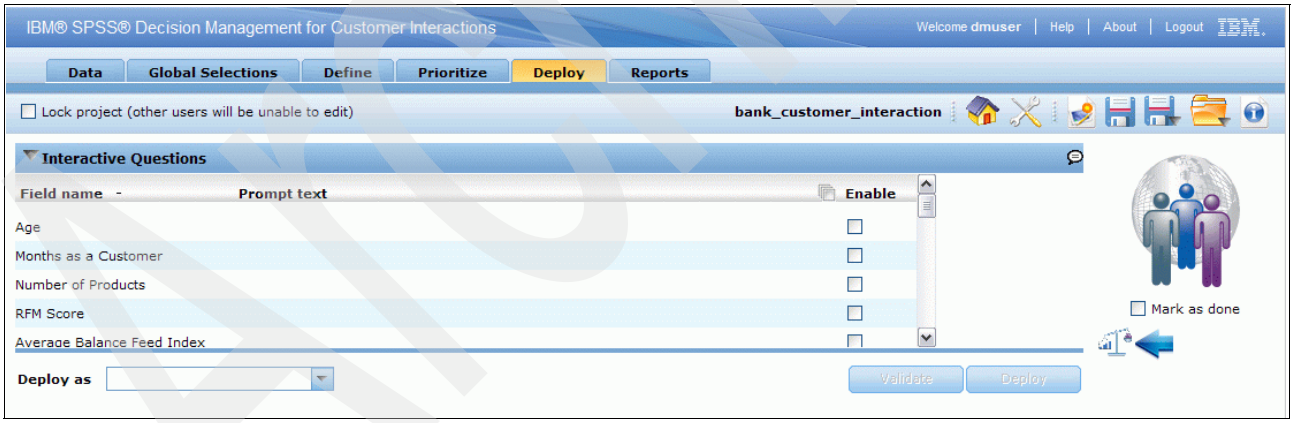

Figure 28 Deploying to production within SPSS Decision Management

SPSS Decision Management can help organizations make the best decisions in real time. Figure 29 shows an example of a financial service company that wants to make marketing recommendations when customers contact the company through the call center, branch, and website. Decision Management helps the agent choose the best offer for each customer based on expected profit and balance the needs of retention and cross-sell campaigns.

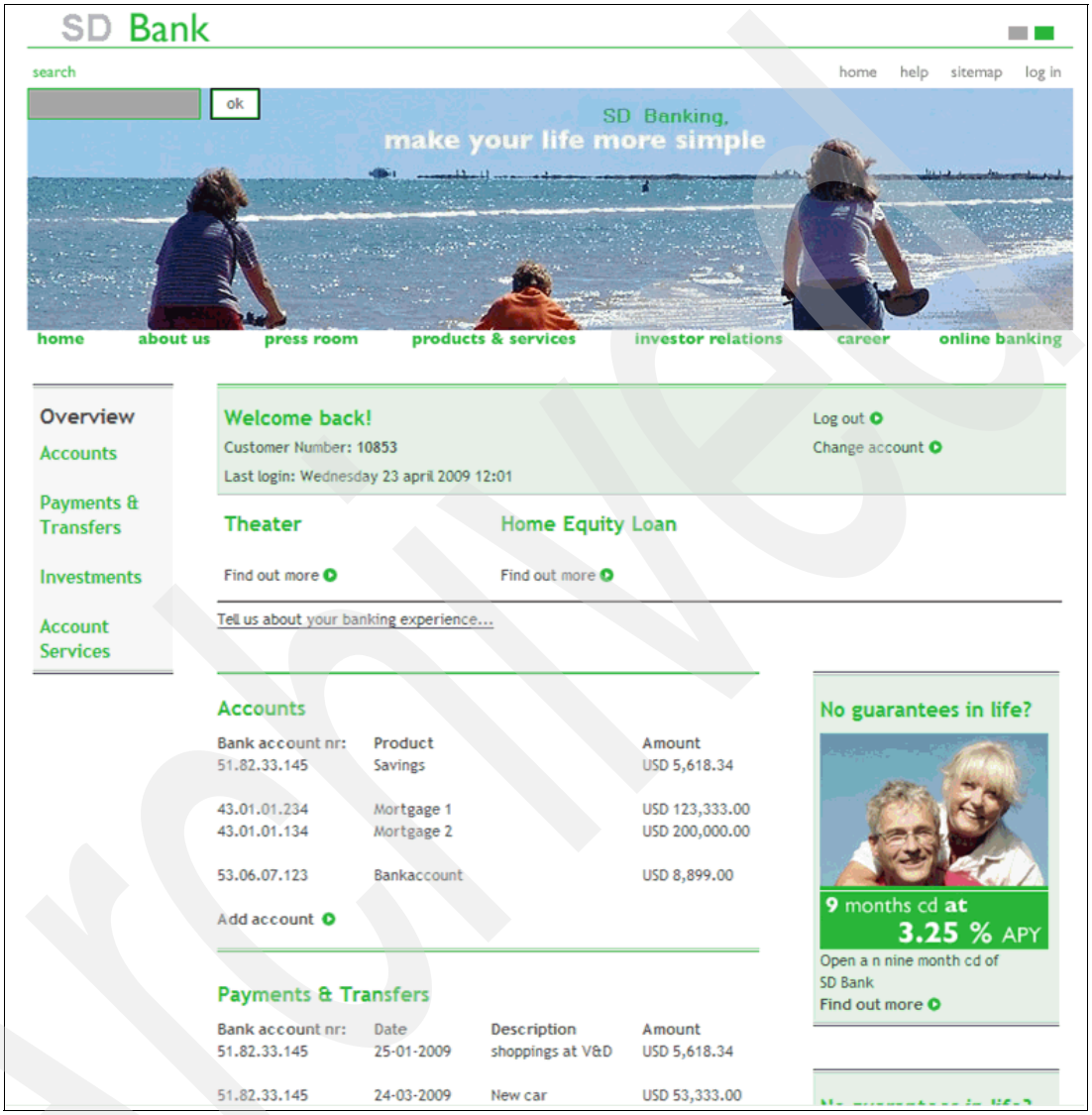

Figure 29 Customer interaction management application processes incoming interactions in real time

Collaboration and Deployment Services product overview

IBM SPSS Collaboration and Deployment Services can help an organization to deploy analytical resources more effectively, applying them in the context of the broader IT stack where they can make the biggest impact.

The Collaboration and Deployment Services platform is a flexible enterprise-level foundation for managing and deploying analytics. It has three key functions:

- Collaboration

  Collaboration is the key to developing and implementing analytics throughout the enterprise. IBM SPSS Collaboration and Deployment Services enables the sharing and reusing of assets efficiently, protecting them in ways that meet internal and external compliance requirements, and publishing results so that a greater number of business users can view and interact with results.

- Automation

  Automation enables your organization to make analytics a core component of daily decision-making processes. You can construct flexible analytical processes that can be operationalized, ensuring consistent results. You also have the tools you need to govern analytical environments just as you manage other business processes.

- Deployment

  Deployment bridges the gap between analytics and action by enabling organizations to operationalize analytics, which means embedding analytic results in front-line business processes. Linkage between IBM Collaboration and Deployment Services and other IBM SPSS products, such as Decision Management, supports the automation of recommendations and tactical decisions. Integration with your existing infrastructure, using standard programming tools and interfaces, makes combining these technologies efficient for IT to deploy and manage and also eases the learning curve for users.

Collaboration and Deployment Services highlights

Key benefits of using IBM SPSS Collaboration and Deployment Services are as follows:

- Bring control to analytical processes by centralizing and automating the evaluation and deployment of models.
- Enhance model accuracy through champion and challenger testing.
- Deploy scores generated by models in real time to support decision-making.
- Integrate analytics within key business processes.

Sample images

This section shows several sample images of the IBM SPSS Collaboration and Deployment Services.

The IBM SPSS Collaboration and Deployment Services, shown in Figure 30, is a flexible, enterprise-level platform for managing and deploying analytics.

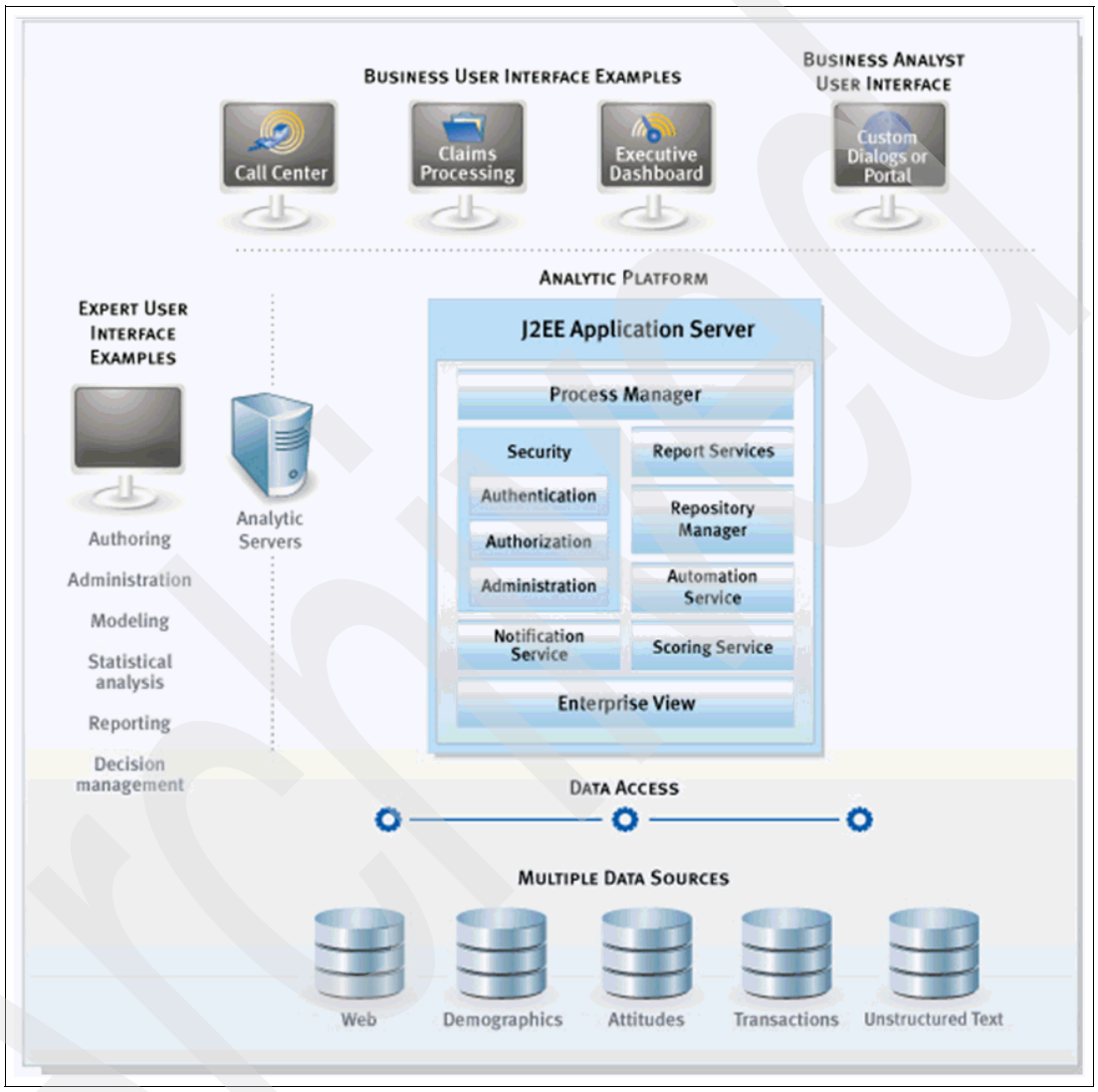

Figure 30 Collaboration and Deployment Services

SPSS Collaboration and Deployment Services enables users to share analytic assets during development, protect them in a secure repository and publish them for consumption by business analysts and users. See Figure 31.

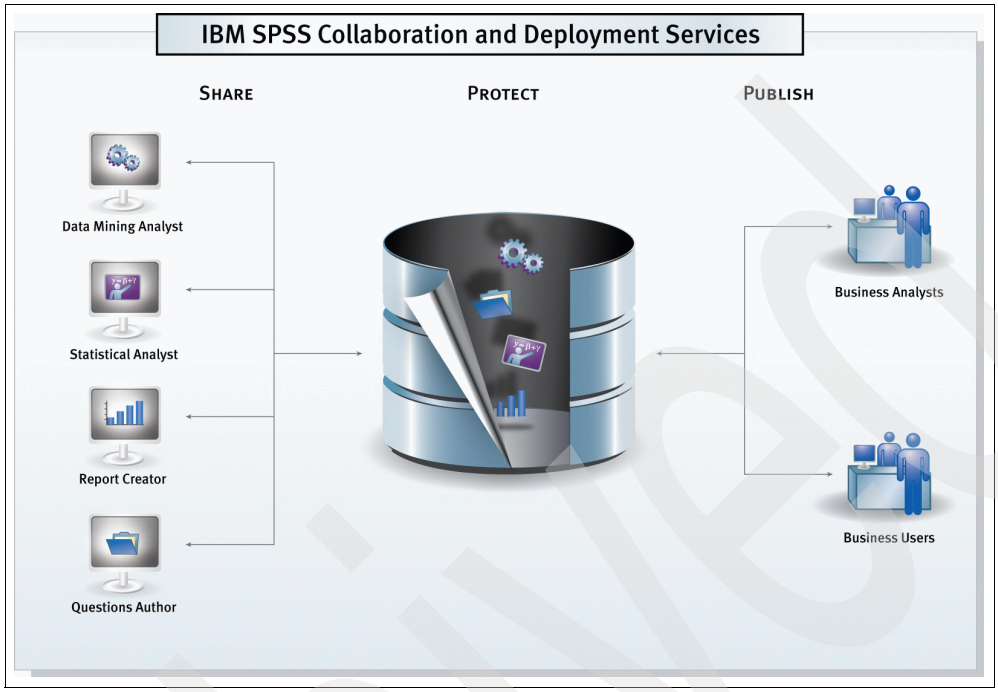

Figure 31 SPSS Collaboration and Deployment Services secure repository

Figure 32 shows an example job flow within SPSS Collaboration and Deployment Services.

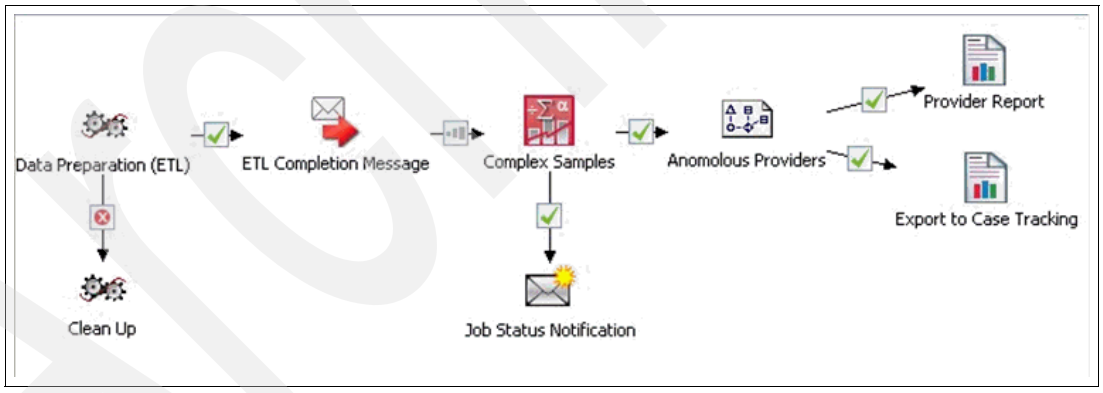

Figure 32 SPSS Collaboration and Deployment Services example job flow

Figure 33 shows the output view in Deployment Manager within SPSS Collaboration and Deployment Services.

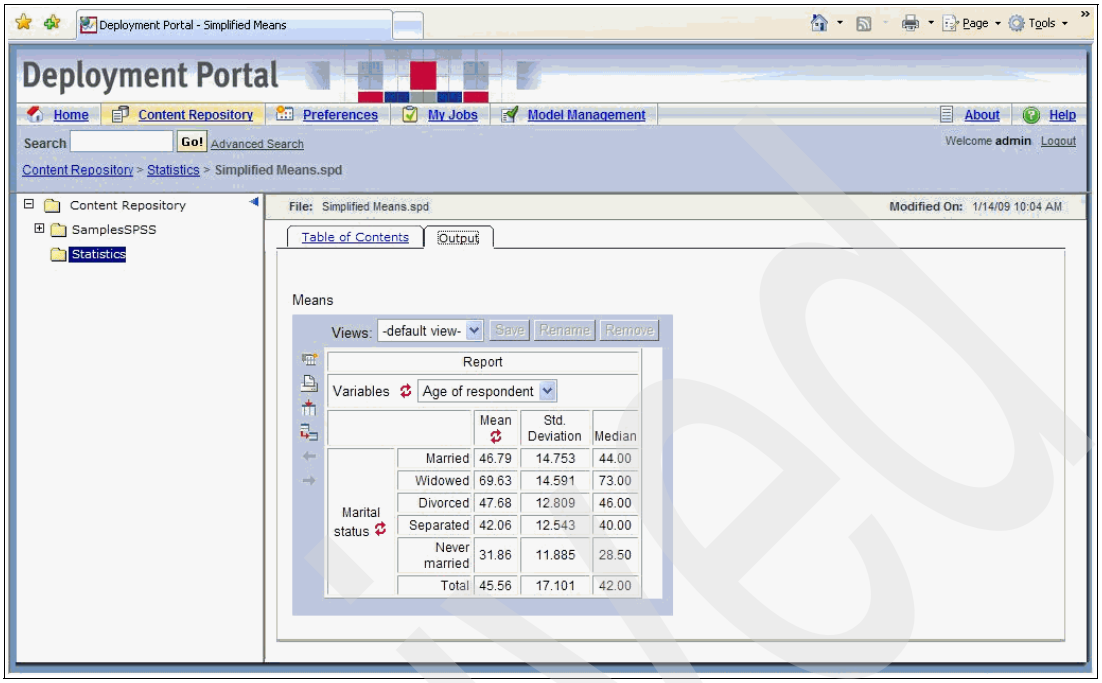

Figure 33 SPSS Collaboration and Deployment Services output view

Running a job using SPSS Collaboration and Deployment Services is shown in Figure 34.

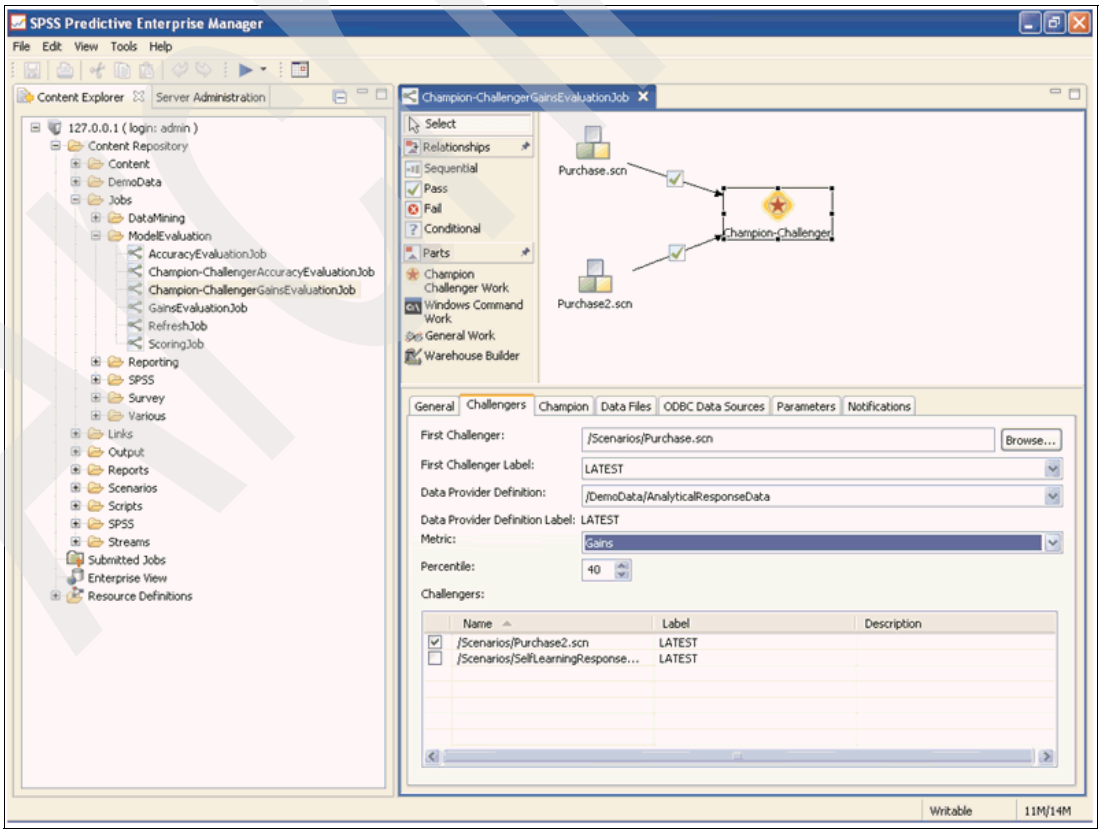

Figure 34 Running a job using SPSS Collaboration and Deployment Services

IBM SPSS reference architecture

This section examines two IBM SPSS reference architecture concepts:

- Predictive enterprise conceptual architecture
- Analytical topology (logical architecture)

Predictive enterprise conceptual architecture

A *predictive enterprise* is defined as an organization that is fully enabled, throughout all areas of its business, by predictive analytics. This concept is based on opportunities to optimize the *moments of truth* of an organization's typical interaction with a customer or constituent.

The *predictive enterprise architecture* aligns with the *Capture* → *Predict* → *Act* phases of the analytical process, described in "The analytical process" on page 23. Data is captured through various channels (during *moments of truth*), and the lines of business (LOB) use predictive insights to manage these moments in ways that address organizational goals. Examples include attracting profitable customers, growing customer value, retaining the best customers, detecting and preventing fraud, and mitigating risk. The overall predictive enterprise conceptual architecture is shown in Figure 35.

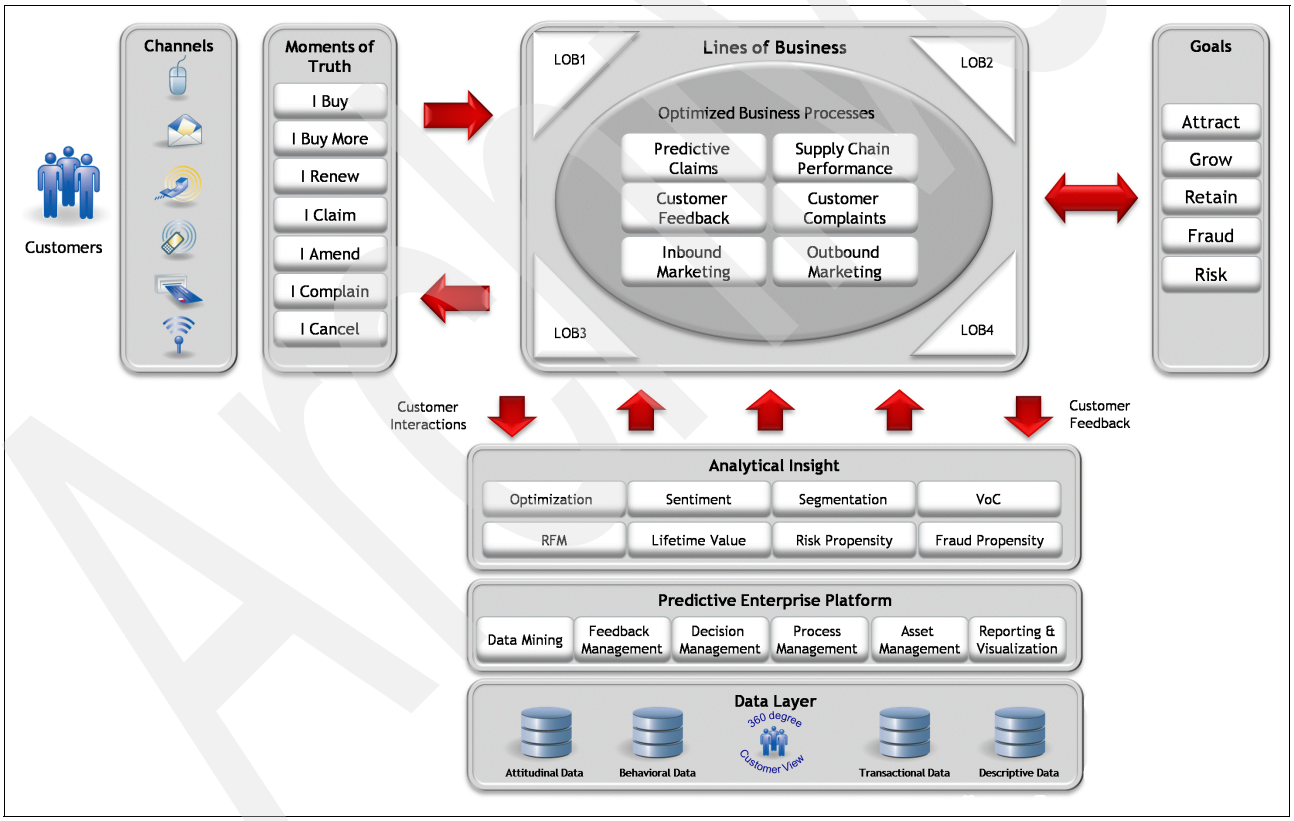

Figure 35 Conceptual architecture

Analytical topology (logical architecture)

The *analytical topology* positions the customer *touch points* in the *Presentation Layer* of the logical architecture (Figure 36 on page 46), or the point at which interaction happens. This

approach can be accomplished through a variety of channels such as postal mail, in-person service, or web browsing.

The *Business Logic Layer* of the architecture provides the analytical insights. These functions are not visible to the beneficiaries of the improved interactions; rather the functions drive the quality of them by bridging analytics and actionable outcomes. An example is understanding a propensity score so that the most effective interaction at the Presentation Layer can be accomplished.

The *Analytics Layer* combined with the system layer of an organization forms the *predictive analytics platform*. This is the core of the predictive capabilities and includes capturing multiple types of data and applying predictive models or statistical analysis. Supporting technologies such as business intelligence (BI) provides the ability to deploy these insights on the operational level so the outcomes from the predictive capabilities can move to the Business Logic phase.

The *Data Layer* is sources of four critical types of data: *descriptive*, *interaction*, *behavioral*, and *attitudinal*. The presence and richness of all four types of data largely influences the breadth of the predictive analytic solution and the downstream ability to improve outcomes at the Business Logic and Presentation Layers.

Given the importance of comprehensive data in the entire predictive process, both structured and unstructured data are critical components used by the IBM SPSS predictive capabilities in developing a predictive enterprise.

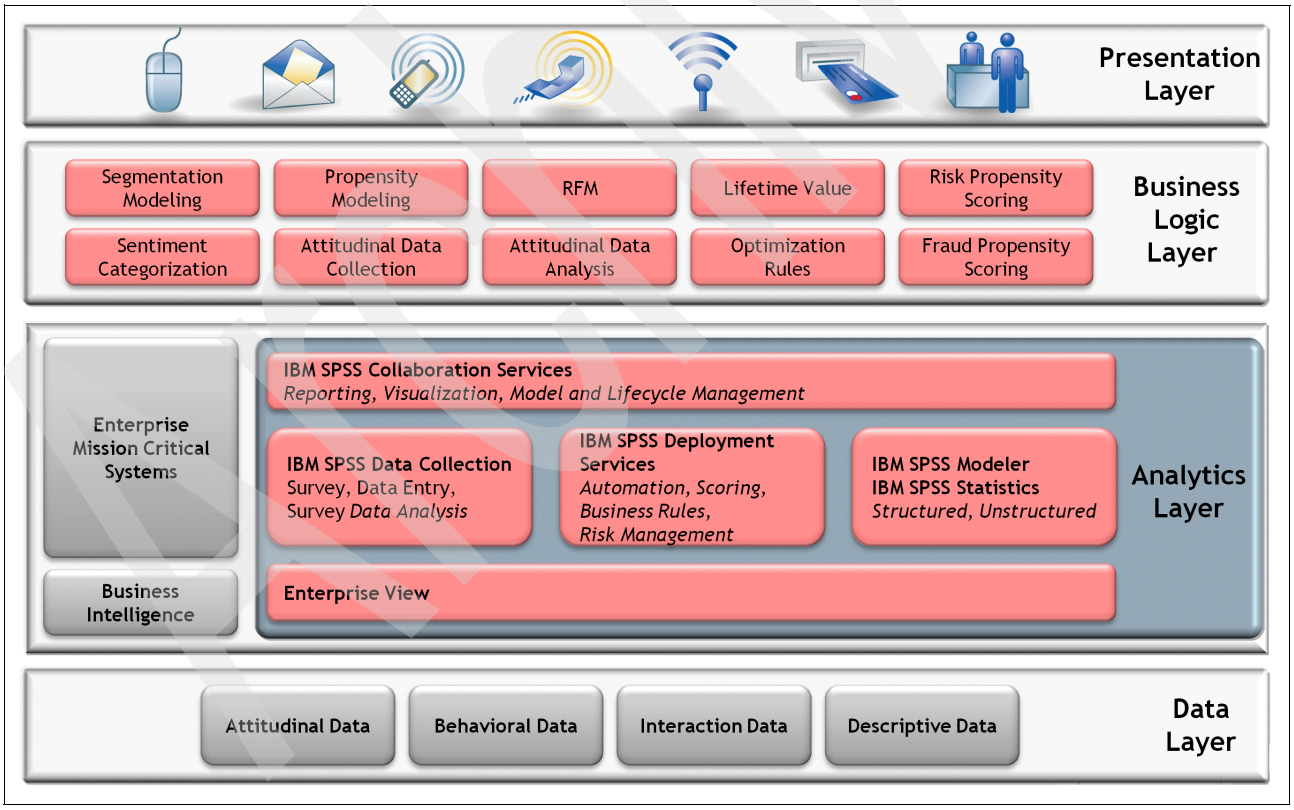

Figure 36 Logical architecture

IBM SPSS on System z

The IBM System z® platform is specifically designed to provide a high performance, highly scalable, and extremely reliable infrastructure that simplifies operational complexity by enabling organizations to deploy a number of critical software applications on a single platform.

The virtualization services, quality of service, and high performance capabilities of System z allows organizations to perform the following tasks:

- Run a complete solution suite on a single piece of hardware, rather than on a room full of servers or on dozens of distributed servers scattered throughout the enterprise.
- Manage virtual servers with a common management interface.
- Host hundreds or thousands of virtual servers within a single footprint with high quality of service provided by the System z environment's failover capabilities.

IBM SPSS predictive analytic software on System z includes the following items:

- IBM SPSS Modeler for Linux® on System z
- IBM SPSS Statistics for Linux on System z
- IBM SPSS Collaboration and Deployment Services for Linux on System z

Figure 37 is an overview of IBM Business Analytics and Data Warehousing on System z capabilities.

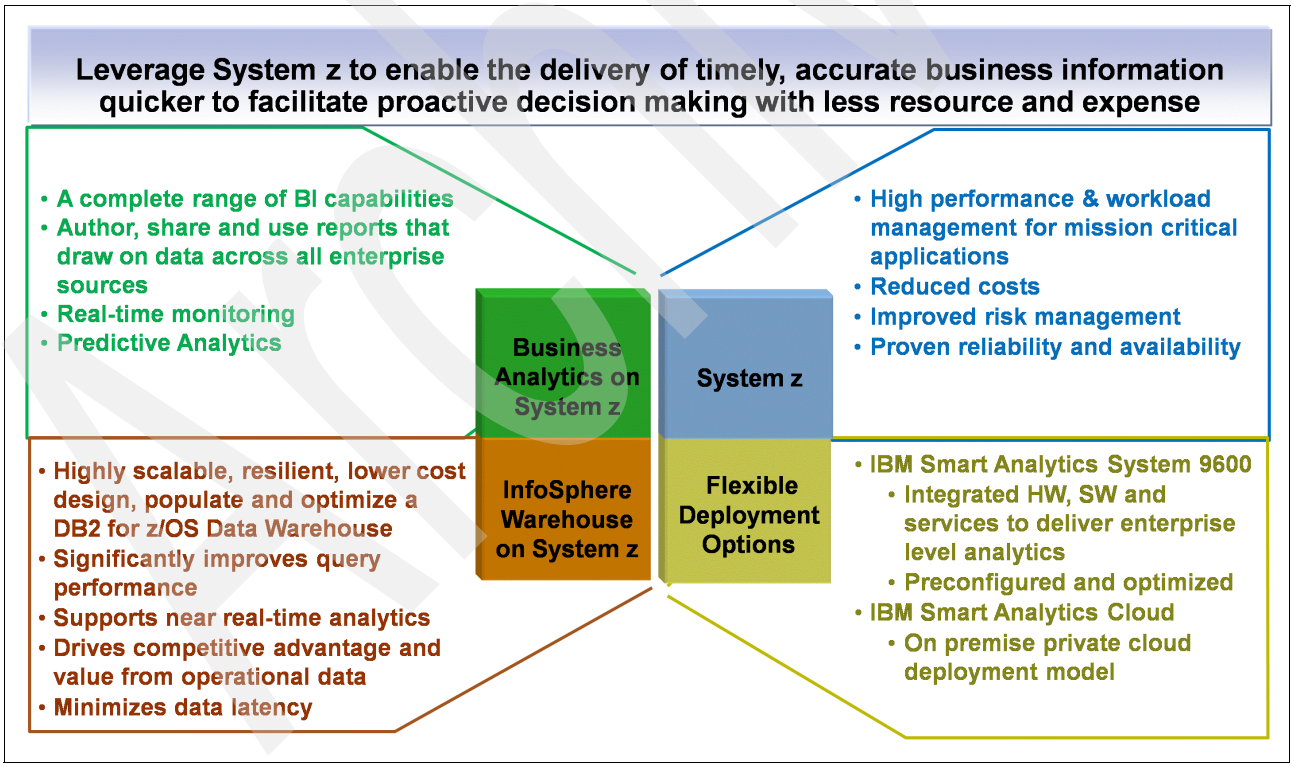

Figure 37 Business analytics and data warehousing on System z

Appendix A: Comparison of IBM SPSS Modeler editions

IBM SPSS Modeler Professional and IBM SPSS Modeler Premium are available in both a desktop-based client deployment and a client/server deployment model. The features listed in Table 1 are accessed from the client. IBM SPSS Modeler Server, available as both a Professional and Premium edition, provides server-based processing and performance enhancement, and additional features such as batch processing, SQL push-back and in-database mining.

Table 1 IBM SPSS Modeler editions feature comparison

Features	Modeler Professional	Modeler Premium
Create a wide range of interactive graphs with automatic assistance	X	X
Use visual link analysis to see the associations in your data	X	X
Interact with data by selecting regions or items on a graph and view the selected information; or select key data for use in analysis	X	X
Access IBM SPSS Statistics tools directly from interface	X	X
Read from and write to operational data from a variety of operational data sources such as IBM DB2®, Oracle, Microsoft® SQL Server, Informix®, Neoview, Netezza, mySQL (Sun) and Teradata.	X	X
Read from and write to data views and metadata stored in Cognos® 8 Business Intelligence	X	X
Import from and export to delimited and fixed-width text files, any IBM SPSS Statistics file, SAS, IBM Data Collection data sources, or XML	X	X
Use data-cleaning options that remove or replace invalid data, automatically impute missing values and mitigate for outliers and extremes	X	X
Apply automatic data preparation to interrogate and condition data for analysis in a single step	X	X
Use field filtering, naming, derivation, binning, re-categorization, value replacement, and field reordering	X	X
Use record selection, sampling (including clustered and stratified sampling), merging (including inner joins, full outer joins, partial outer joins, and anti-joins), and concatenation; sorting, aggregation, and balancing	X	X
Use data restructuring, partitioning and transposition	X	X
Extensive string functions: string creation, substitution, search and matching, white space removal, and truncation	X	X
RFM scoring: aggregate customer transactions to provide Recency, Frequency, and Monetary value scores and combine these to produce a complete RFM analysis.	X	X
Use interactive model and equation browsers and view advanced statistical output	X	X
Show relative effect of various data attributes on predicted outcomes with variable importance graphs	X	X
Combine multiple models or use one model to analyze a second model	X	X
Use automatic classification (both binary and numeric) and clustering models to eliminate need for selecting individual algorithm	X	X

Features	Modeler Professional	Modeler Premium
Use Modeler's Component-Level Extension Framework (CLEF) to build custom algorithms	X	X
Through the integration of IBM SPSS Statistics, use R language to extend analysis options through customization	X	X
C&RT, CHAID, QUEST: Decision-tree algorithms including interactive tree-building	X	X
Decision List: Interactive rule-building algorithm	X	X
K-Means, Kohonen, Two Step, Discriminant, Support Vector Machine (SVM): Clustering and segmentation algorithms	X	X
Generalized rule induction (GRI) association discovery algorithm	X	X
Factor/PCA, Feature Selection: Data reduction algorithms	X	X
Regression, Linear, GenLin (GLM): Linear equation modeling	X	X
Self-learning response model (SLRM): Bayesian model with incremental learning	X	X
Time-series: Generate and automatically select time-series forecasting models	X	X
C5.0 decision tree and rule set algorithm	X	X
Neural Networks: Multi-layer perceptrons with back-propagation learning, and radial basis function networks	X	X
Support Vector Machines: Advanced algorithm with accurate performance for wide data sets	X	X
Bayesian Networks: Graphical probabilistic models	X	X
Cox regression: Calculate likely time to an event	X	X
Anomaly Detection: Detect unusual records through the use of a cluster-based algorithm	X	X
KNN: Nearest neighbor modeling and scoring algorithm	X	X
Apriori: Popular association discovery algorithm with advanced evaluation functions	X	X
CARMA: Association algorithm which supports multiple consequents	X	X
Sequence: Sequential association algorithm for order-sensitive analyses	X	X
Export models using SQL or PMML (the XML-based standard format for predictive models)	X	X
Extract text data from files, operational databases and RSS feeds (that is blogs, web feeds)		X
Select native language extractor options for Dutch, English, French, German, Italian, Portuguese, Spanish, or Japanese, or translate virtually any language using Language Weaver (separately licensed)		X
Extract domain-specific concepts such as uniterms, expressions, abbreviations, acronyms, and more		X
Calculate synonyms using sophisticated linguistic algorithms and embedded or user-specified linguistic resources		X

Features	Modeler Professional	Modeler Premium
Name concepts by person, organization, term, product, location, and other user-defined types		X
Extract non-linguistic entities such as address, currency, time, phone number, and social security number (SSN)		X
Use and customize pre-built templates and libraries for sentiment analysis, CRM, security and intelligence, market intelligence, life sciences, and IT		X
Use pre-packaged Text Analytics Packages (TAPs) for the most common business applications and create your own		X
Create clusters based on term co-occurrence using concept clustering algorithms, which provide an at-a-glance view of main topics and the way in which they are related		X
Intelligently group text documents and records based on content, using text classification algorithms -Enable advanced concept selection and deselection for use in predictive modeling		X
Use text-based and visual reports to interrogate concept relationship, occurrence, frequency and type		X
Identify and extract sentiments (for example, likes and dislikes) from text in Dutch, English, French, German, and Spanish		X
Read from IBM Classic Federation data sources	X	X

Appendix B: Comparison of IBM SPSS Statistics editions

Table 2 compares the feature sets of the IBM SPSS Statistics packages.

Table 2 IBM SPSS Statistics editions feature comparison

Features	IBM SPSS Statistics Standard	IBM SPSS Statistics Professional	IBM SPSS Statistics Premium
Core statistical and graphics capabilities to take standard analytic projects from start to finish	X	X	X
Seamless integration with R, Python and other environments to easily and effectively expand statistical capabilities and programmability	X	X	X
Advanced statistical procedures, including GLM, GLMM, HLM, GENLIN & GEE, to more accurately identify and analyze complex relationships	X	X	X
Nonlinear regression, including MLR, Binary Logistic Regression, NLR, CNLR & Probit Analysis, to improve the accuracy of predictions	X	X	X
Customized tables to analyze and report on numerical and categorical data	X	X	X
Missing value analysis (with imputation) to address issues of “dirty data” for more complete analysis and better decision-making		X	X
Advanced data preparation to identify anomolies and the other data that can skew results		X	X
Decision trees to better identify groups, discover relationships between groups, and predict future events.		X	X
Forecasting to predict trends and build expert time-series forecasts		X	X
Structural Equation Modeling to test hypotheses and confirm relationships among observed and latent variables			X
Bootstrapping to test the stability and reliability of predictive models			X
Advanced sampling assessment and testing procedures			X
High-end charts and graphs to aid analysis and reporting			X
Direct marketing and product decision-making procedures to identify best customers and the product attributes that appeal to them			X
Add Statistics Server to this package and realize dramatically faster performance, improved productivity and powerful automation	X	X	X

Other products that complement IBM SPSS Statistics are as follows:

- IBM SPSS Amos
- IBM SPSS Advantage for Microsoft Excel
- IBM SPSS SamplePower
- IBM SPSS Statistics Developer
- IBM SPSS Statistics Server
- IBM SPSS Text Analytics for Surveys
- IBM SPSS Visualization Designer

Summary

This document provides an overview of the IBM SPSS Predictive Analytics capabilities, the IBM SPSS Predictive Analytics software suite, and a discussion of real-world applications of predictive analytics solutions within various industries: banking, insurance, retail, telecommunications, public sector, and industrial.

The IBM commitment to helping organizations solve real business problems with predictive analytics includes a focus on empowering everyone in the enterprise to use the value of data-driven decision-making at the point of impact and improve outcomes. This is a critical source of advantage over traditional methods for managing business objectives, which often limited the use of analytics to highly specialized back-office functions.

IBM SPSS Predictive Analytics product portfolio includes a comprehensive set of end-to-end advanced mathematical and statistical capabilities to extract predictive knowledge that, when deployed into existing processes, makes them adaptive to improve outcomes.

IBM SPSS Predictive Analytics software enables organizations to capture, predict, act:

- *Capture* complete information about people's attitudes and opinions.
- *Predict* the outcomes of interactions before they occur.
- *Act* on predictive insights by embedding analytic results into business processes.

It is one of the most comprehensive and powerful predictive analytics suites available today.

Other resources for more information

See the following resources for additional information:

- IBM SPSS software overview:

  http://www.spss.com/software/?source=homepage&hpzone=nav_bar
- Nucleus Research, 2005, Research Note F31, *The Real ROI from SPSS*:

  http://www.spss.com/media/collateral/solutions/Nucleus_Research-ROI_From_SPSS.pdf
- IBM SPSS white papers:

  http://www.spss.com/downloads/
- IBM Business Analytics website:

  http://www.ibm.com/software/analytics/
- IBM white paper *Building a Smarter Planet with IBM System z:*

  ftp://ftp.software.ibm.com/common/ssi/pm/br/n/zsb03033usen/ZSB03033USEN.PDF

The team who wrote this guide

This guide was produced by a team of specialists from around the world working at the International Technical Support Organization, Austin Center.

Theresa Morelli is the IBM SPSS Worldwide Industry Solutions team's Marketing Manager and is responsible for developing core messaging and go-to-market strategy for IBM SPSS Predictive Analytics capability within the Business Analytics division of the IBM Software Group. Prior to joining SPSS in 2007, Theresa spent ten years managing various teams within Accenture's CIO organization, led the design and implementation of the support model for Accenture's SAP solution, and defined the release management process that governs the delivery of activities necessary for the operation of Accenture's 800+ internal and customer applications. Theresa holds a Master of Business Administration degree from The University of Chicago Booth School of Business with concentrations in marketing, strategy, and organizational behavior.

Colin Shearer is the SPSS Worldwide Industry Solutions Leader. With a background in computer science and artificial intelligence at the University of Aberdeen, he has been involved since 1984 in applying advanced software solutions to business problems.

Previously with SD-Scicon and Quintec Systems, he was one of the founders of Integral Solutions Ltd. (ISL) in 1989. A pioneer of data mining in the early 1990s, he was the architect of ISL's award-winning Clementine system (now IBM SPSS Modeler), and led a team that tackled numerous successful data mining applications in areas including finance, broadcasting, market research, and defense. In December 1998, SPSS acquired ISL, and Shearer became responsible for a worldwide team of data mining consultants and for SPSS Advanced Data Mining Group. He held various positions at SPSS including global head of Product Marketing and Senior Vice President for Market Strategy; following the acquisition of SPSS by IBM, he moved to his current position in January 2010.

Axel Buecker is a Certified Consulting Software IT Specialist at the International Technical Support Organization (ITSO), Austin Center. He writes extensively and teaches IBM classes worldwide in areas of software security architecture and network computing technologies. He holds a degree in computer science from the University of Bremen, Germany. He has 24 years of experience in a variety of areas related to workstation and systems management, network computing, and e-business solutions. Before joining the ITSO in March 2000, Axel worked for IBM in Germany as a Senior IT Specialist in Software Security Architecture.

Special acknowledgement to the following subject matter experts who developed industry content that is included in this guide:

Martijn Wiertz, Director, SPSS Industry Solutions Team & WW Insurance Solutions Leader

Keith Ellis, Worldwide Predictive Analytics Banking Solution Leader

Colin Linsky, Worldwide Predictive Analytics Retail Solution Leader

Eric Martin, Worldwide Predictive Analytics Telco Solution Leader

Scott Mutchler, Worldwide Predictive Analytics Industrial Solution Leader

Ken Stehlik-Barry, Worldwide Predictive Analytics Education Solution Leader

Jaap Vink, Worldwide Predictive Analytics Public Sector Solution Leader

Daniel Ng, Director, Enterprise Architecture

Kathy Konkel, Manager Indirect Channels - SPSS Product Marketing

Erik Dingvall, EMEA Enterprise Architecture Manager

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks
- Follow us on Twitter:

  http://twitter.com/ibmredbooks
- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806
- Explore new IBM Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm
- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document, REDP-4710-00, was created or updated on November 12, 2010.

Trademarks