The Analytics Revolution: Optimizing Reporting and Analytics to Make Actionable Intelligence Pervasive

Prepared for IBM by:

David Loshin
Knowledge Integrity, Inc.
January, 2010

Growing Data Volumes and Smarter Decisions

For years, technical analysts have been speculating about the monumental growth of data. A 2008 report suggests that data volume will continue to expand at a healthy rate, with "the size of the largest data warehouse … *triples approximately every two years.*"[1] And what seemed to be huge just a few years ago seems almost puny today: that same report mentions "a large commercial entity – that will soon have about *300 TB (terabytes)* of data in its enterprise warehouse," a size that does not seem to be terribly outrageous. That being said, structured data is no longer the key culprit – its growth is "far outpaced by a 61.7% CAGR predicted for *unstructured data* in traditional data centers."[2]

Complex business processes are increasingly executed through a variety of interconnected systems, and integrated sensors and probes not only enable continuous measurement of operational performance, system interconnectedness allows for rapid communication and persistence of those measures. Every day, it is estimated that 15 petabytes of new information is being generated, 80% of which is unstructured[3].

But as the volume of data grows, so does the cost of finding those critical pieces of information necessary to make those business processes run at their optimized level. The issue is no longer the need to capture, store, and manage that data. Rather, the challenge is the need for distilling out and delivering the relevant pieces of knowledge to the right people at the right to time to enhance the millions of opportunities for decision-making that occur on a daily basis.

In essence, this can be summarized as the desire to integrate actionable intelligence in a pervasive manner into both the strategic and the operational processes across the management hierarchy. And whether this means notifying senior management of emerging revenue opportunities, providing real-time insight into corporate performance indicators, or hourly realignment of field repair team schedules to best address customer service outages, the ability to accumulate, transform, and analyze information to provide rapid, trustworthy analyses to the right people at the right time can add significant benefits to growth and competitiveness.

Delivering Actionable Intelligence

When massive amounts of unfiltered data are channeled across the organization, overwhelmed individuals are stunned into "analysis paralysis" – the compulsion to delay decision-making while waiting for just a *little bit more data* that can simplify (and perhaps justify) that impending decision. This paralysis can be alleviated when the information overload is throttled back to filter out the specific information necessary to optimally drive the decision process and allow specific actions to be taken.

[1]Richard Winter, "Why Are Data Warehouses Growing So Fast?" – B-eye-network.com, April 2008
[2]Beth Pariseau, "IDC: Unstructured data will become the primary task for storage," reference to IDC survey at techtarget.com, October 2008
[3] Bates, Pat, Biere, Mike, Weideranders, Rex , Meyer, Alan, and Wong, Bill, "New Intelligence for a Smarter Planet," ftp://ftp.software.ibm.com/common/ssi/pm/bk/n/imm14055usen/IMM14055USEN.PDF

Delivering trustworthy **actionable intelligence** to the right people when they need it short-circuits analysis paralysis and encourages rational and confident decisions.

At one end of the continuum, overall company performance is reviewed and alternatives are considered to help adjust corporate strategy for long-term value generation. At the other end, operational activities are improved with specific pieces of intelligence that can adjust and optimize activities in real time. Of course, actionable intelligence informs both strategic and operational processes, and its pervasive delivery to staff members up and down the "org chart" can facilitate a transition from reacting to what has happened in the past to better predict the optimal choices for the future.

Strategies that Drive Organizational Optimization

As more strategists within the organization are becoming aware of the capabilities of reporting and analysis, there is a growing desire to employ data analysis to proactively manage and improve business strategies. Observing the environment provides insight into how the organization works, as well as allows the analyst to understand areas of failure, success, and differentiate those tactics that can be replicated and thereby lead to greater success.

Traditional reporting within a longitudinal context allows the analyst to see what has happened within organization and review trending over time. Organizing the data along different quantitative and qualitative hierarchical dimensions such as location, organization, customer profile, product categories, among many others, lets the analysts slice and dice the data to look for explicit business opportunities or process improvements. Interactive environments such as dashboards and mash-ups facilitate informed strategic decisions.

Upon making adjustments to the strategic direction of the organization, the incorporation of real-time delivery of key performance metrics allows senior stakeholders to review the impact and results of those strategic decisions within a rapid turnaround time, thereby increasing agility, reducing risk, and allowing the organization to rapidly respond to emerging business opportunities. More pointedly, gaining a deeper insight into how the organization works (or doesn't work!) can inform strategists to make significant changes to the way the company does business.

Integrating Analytics and Operations

Actually, the results of predictive analytics formulated as actionable intelligence can be integrated directly into operational processes even without the awareness of business consumers. For example, merging the results of predictive analytics with customer profiles empowers decision-makers at all levels of the organization to recognize and react rapidly to emerging opportunities, such as real-time adjustments to call-center scripts, or rerouting of product deliveries based on real-time warehouse and stock room inventory data, point-of-sales data, or even traffic or weather data. Customer preferences, profiles, and web statistics may drive dynamic content realignment on web sites to improve end-user response.

Industry Examples

From one standpoint, opportunities for improvement manifest themselves differently depending on the industry, while from another standpoint there are common dimensions of operations that can be improved no matter which industry. Certainly, a company within a particular industry can benefit from reporting and analytics associated with the specific aspects of that industry, such as these "vertical" examples:

- **Health Care** – Monitoring business process performance permeates all aspects of quality of care. For example, understanding why some practitioners are more successful at treating certain conditions can lead to improved quality of care. Analytics can help to discover the factors that contribute to success of one approach over others, and see whether those successes are dependent on variables within the control of the practitioner or factors outside their control. Improved diagnostic approaches can reduce the demand for high-cost diagnostic resources such as imaging machinery, and better treatments can reduce the duration of patient stays, freeing up beds, improving throughput, and enabling more efficient bed utilization.

- **Logistics/Supply Chain** – Integrated analysis for transportation and logistics management sheds insight into evaluation of many aspects of an efficient supply chain. For example, business intelligence is used to analyze usage patterns for particular products based on a series of geographic, demographic, and psychographic dimensions. Predictability becomes the magic word – knowing what types of individuals in which types of areas account for purchases of the range of products over particular time periods can help in more accurately predicting (and therefore meeting) demand. As a result, the manufacturer can route the right amounts of products to reduce or eliminate out-of-stocks. At the same time, understanding demand by region over different time periods leads to more accurate planning of delivery packaging, methods, and scheduling. One can map the sales of products in relation to distance from the origination point; if sales are lower in some locations than others, it may indicate a failure in the supply chain that can be reviewed and potentially remediated in real time.

- **Telecommunications** – In an industry continually battling customer attrition, increasing a customer's business commitment contributes to maintaining a long customer lifetime. For example, examining customer cell phone usage can help to identify each individual's core network. If a customer calls a small number of residential land lines or personal mobile phones, that customer may be better served by a "friends and family" service plan that lowers the cost for the most frequently called numbers. Identifying household relationships within the core network may enable service bundling, either by consolidating mobile accounts, or by cross-selling additional services such as landline service, internet, and other entertainment services. On the other hand, if the calls from the customer's individual mobile phone are largely to business telephone numbers and have durations between a half hour to an hour, that customer may be better served with a business telephony relationship that bundles calling with additional mobile connectivity services.

- **Retail** – The large volume of point of sales data makes it a ripe resource for analysis, and retail establishments are always looking for ways to optimize their product placement to increase sales while reducing overhead to increase their margins, especially when market baskets can be directly tied to individuals via affinity cards. Understanding the relationship between a brick-and-mortar store location and the types of people that live within the surrounding area helps the store managers with their selection of products for store assortment. Strategic product placement (such as middle shelf or end-cap) can be reserved for those items that drive profitability, and this can be based on a combination of product sales by customer segment coupled with maps of customer travel patterns through the store. Product placement is not limited to physical locations; massive web logs can be analyzed for customer behavior to help dynamically rearrange offer placement on a web site, as well as encourage product upselling based on abandoned cart analysis, through collaborative filtering, or based on the customer's own preferences.

- **Financial Services/Insurance** – In both insurance and banking, identifying risks and managing exposure are critical to improved profitability. Banks providing a collection of financial services develop precise models associated with customer activities and profiles that identify additional risk variables. For example, analyzing large populations of credit card purchases in relation to mortgage failures may show increased default risk for individuals shopping at particular shopping malls or eating at certain types of fast food restaurants. In turn, recognizing behaviors that are indicative of default risk may help the bank anticipate default events and reach out to those individuals with alternate products that keep them in their homes, reduce the risk of default, and improve predictability of the loan's cash flow over long periods of time.

- **Manufacturing -**Plant performance analysis is critical to maintaining predictable and reliable productivity; tracking production line performance, machinery downtime, production quality, work in progress, safety incidents, and delivering measurements of operational performance indicators along the management escalation chain so that adverse events can be addressed within the proper context within a reasonable timeframe.

- **Hospitality** – Hotel chains assess customer profiles and related travel patterns ,and know that certain customers may be dividing their annual "night allocation" among the competitors. By analyzing customer travel preferences and preferred locations, the company may present incentive offers through the loyalty program to capture more of that customer's night allocation.

The examples for these industries are similar in that the analysis ranges from straightforward reporting of key business performance indicators to exploring opportunities for optimizing the way the organization is run or improving interactions with customers and other business partners. Other industries benefit from reporting and analytics.

Common Value Drivers

Given the common desire to continuously improve business interactions with customers and partners, we can consider analytics related to common value drivers that expose opportunities for optimization applied across all industries. For example, all businesses must satisfy the needs of a constituent community (most frequently referred to as "customers"), as well as supporting internal activities associated with staff management and productivity, spend analysis, asset management, project management, etc. Consider these examples:

- **Spend Analysis** – Including the collection, standardization, and categorization of product purchase and supplier data to reduce costs, improve the predictability of high-value supply chains, identify fraud, and improve efficiency.

- **Customer Profiling** – This includes customer analytics that encompass the continuous refinement of individual customer profiles that incorporate demographic, psychographic, and behavioral data about each individual to support segmented and micromarketing along numerous relevant taxonomies.

- **Product Price Modeling –** This is used to analyze price points looking for the sweet spot that will encourage the largest number of customers while maximizing profitability.

- **Targeted Marketing** – Given a set of customer likes and dislikes can augment a marketing campaign to target small clusters of customers that share profiles. In fact, laser-style marketing focused directly at individuals as a byproduct of customer analytics.

- **Personalization** – The process of crafting a presentation to the customer based on that customer's profile is the modern-day counterpart to the old-fashioned salesman who remembers everything about his individual "accounts." Business intelligence exploits customer profiles to customize the presentation of material or content for an individual, and is meant to enhance that customer's experience.

- **Customer Satisfaction** – Another benefit of the customer profile is the ability to provide customer information to the customer satisfaction representatives. This can improve the representative's ability to deal with the customer and expedite problem resolutions.

- **Customer Lifetime Value** – How does a company determine who their best customers are? Customer lifetime analytics compute the lifetime value of a customer as a measure of a customer's profitability over the lifetime of the relationship, incorporating the costs associated with managing that relationship as well as the revenues expected from that customer.

Why Pervasive Analytics?

The concept of "business intelligence" and "analytics" include tools and techniques supporting a collection of user communities across an organization, as a result of collecting and organizing numerous

(and diverse) data sets to support both management and decision making at operational, tactical, and strategic levels. Through data collection, aggregation, analysis, and presentation, actionable intelligence can be delivered to best serve a wide range of target users. Organizations that have matured their data warehousing programs allow those users to extract actionable knowledge from the corporate information asset and rapidly realize business value.

But while traditional data warehouse infrastructures support business analyst querying and canned reporting or senior management dashboards, a comprehensive program for information insight and intelligence can enhance decision-making process for all types of staff members in numerous strategic, tactical, and operational roles. Even better, integrating the relevant information within the immediate operational context becomes the differentiating factor. Offline customer analysis providing general sales strategies is one thing, but real-time actionable intelligence can provide specific alternatives to the sales person talking to a specific customer based on that customer's interaction history in ways that best serve the customer while simultaneously optimizing corporate profitability as well as the salesperson's commission. Maximizing overall benefit to all of the parties involved ultimately improves sales, increases customer and employee satisfaction, and improves response rate while reducing the cost of goods sold – a true win-win-win for everyone.

The wide range of analytical capabilities all help suggest answers to a series of increasingly valuable questions:

- **What?** – Predefined reports will provide the answer to the operational managers, detailing what has happened within the organization and various ways of slicing and dicing the results of those queries to understand basic characteristics of business activity (e.g., counts, sums, frequencies, locations, etc.).Traditional BI reporting provides 20/20 hindsight – it tells you what has happened, it may provide aggregate data about what has happened, and it may even direct individuals with specific actions in reaction to what has happened.

- **Why?** – More comprehensive ad hoc querying coupled with review of measurements and metrics within a time series enables more focused review. Drilling down through reported dimensions lets the business client get answers to more pointed questions, such as the sources of any reported issues, or comparing specific performance across relevant dimensions.

- **What if?** – More advanced statistical analysis, data mining models, and forecasting models allow business analysts to consider how different actions and decisions might have impacted the results, enabling new ideas for improving the business.

- **What next?** – By evaluating the different options within forecasting, planning, and predictive models, senior strategists can weigh the possibilities and make strategic decisions.

- **How?** – By considering approaches to organizational performance optimization, the C-level managers can adapt business strategies that change the way the organization does business.

Information analysis makes it possible to answer these questions. Improved decision-making processes depend on supporting business intelligence and analytic capabilities that increase in complexity and value across a broad spectrum for delivering actionable knowledge (as is shown in). As the analytical functionality increases in sophistication, the business client can gain more insight into the mechanics of optimization. Statistical analysis will help in isolating the root causes of any reported issues as well as provide some forecasting capabilities should existing patterns and trends continue without adjustment. Predictive models that capture past patterns help in projecting "what-if" scenarios that guide tactics and strategy towards organizational high performance.

Reports	Ad Hoc	Search	Dashboards	Statistical Analysis	Forecasting Models	Planning	Predictive Models	Optimizing Models

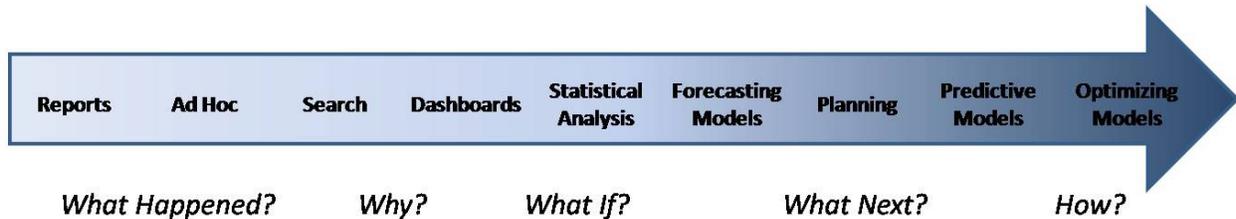*What Happened?* *Why?* *What If?* *What Next?* *How?*

Figure 1: A range of techniques benefits a variety of consumers for analytics.

Intelligent analytics and business intelligence are maturing into tools that can help optimize the business. That is true whether those tools are used to help

- C-level executives review options to meet strategic objectives,

- Senior managers seeking to streamline their lines of business, or

- Operational decision-making in ways never thought possible.

These analytics incorporate data warehousing, data mining, multi dimensional analysis, streams, and mash-ups to provide a penetrating vision that can enable immediate reactions to emerging opportunities while simultaneously allowing one to evaluate the environment over time to see ways to change the business.

Understanding the Technology Considerations

Early data warehousing architects struggled with designing and executing effective approaches for the extraction of data from source systems, transforming that data into a format suitable for analysis, and loading the data into the data warehouse. Once the data has been organized within the data warehouse, reporting and analysis tools such as Online Analytical Processing (OLAP) and query & reporting tools were employed to deliver data to the knowledge consumer.

Although some of the core aspects of those approaches remain unchanged, the details of that technical infrastructure have evolved since the mid 1990's in order to meet the increasing demand for orders of magnitude more data available from many more sources. In turn, both the variety and expectations of the pool of knowledge consumers has increased the demands on the types, scalability, and delivery

mechanisms for actionable knowledge. Today, delivering the right information to the right people at the right time is the culmination of best practices in data management and organization combined with a technical infrastructure designed to direct many steady channels of data into a high performance analysis platform that can deliver trustworthy results within real time constraints. This depends on three areas of technology:

- Continuous synchronization of data from multiple sources to provide a coherent and consistent view;

- Cohesive information integration allowing for massive amounts of high quality data to be harmonized and aligned to enable effective analysis; and a

- Comprehensive set of analytics services that reduce or even eliminate the need to be a power user to derive benefit from actionable intelligence.

In this section we look at some of the key technical considerations necessary for a modern analytics environment, and how the results can be fully integrated into hundreds, if not thousands, of daily business processes. Given an understanding of the technology components, we will begin to see some challenges emerge in environments with heterogeneous technologies cobbled together to support the analytics program.

Continuous Availability: Data Synchrony & Coherence

A recurring challenge of the traditional data warehousing framework involves the elongated turnaround time between the articulated need for reporting or analysis and its delivery. Delaying the delivery of actionable knowledge derived from current data makes it difficult to make timely decisions, and long data latencies will impact the ability to assess the results of those decisions within a reasonable time frame. Essentially, data synchronization has become a critical component within the infrastructure in order to integrate with the strategic and operational decision-making activities. The criticality of timely and current data cannot be understated. For example, increased sales by product category in particular regions may suggest a growing demand requiring immediate reallocation of logistics resources to prevent out-of-stocks and feed that demand. Relying on week-old order information is insufficient; instead, current shelf and warehouse stock information coupled with shipping resource allocation enables immediate decisions that maintain a steady flow of product to the customers.

Pervasive business intelligence and analytics require a high degree of data synchrony, meaning that the environment must

- Reduce or eliminate data latency;

- Maintain coherence of the data being used for analysis across the enterprise;

- Provide timely and current information; and

- Provide consistent, deterministic results to similar requests.

Maintaining a reasonable degree of synchrony and coherence among the multitude of data sources available within (as well as from outside) the organization requires technical strategies for continuous data availability that do not impose a strain on the environment. Some of those strategies include:

- Change data capture (CDC) for data replication, involving a managed, synchronized copying of data from a source to one or more target data systems. CDC is an event-driven mechanism for capturing changes from source datasets and propagating those changes through various channels, either directly to target databases or through a message queue for subsequent processing. *Synchronizing data modifications via CDC enables coherence between operational systems and analytical systems, enabling the discovery of actionable opportunities in real-time while maintaining consistency across reporting systems.*

- Data Federation, which enables transparent access to heterogeneous (and generally physically distributed) data types, platforms, and sources, and in numerous formats, without requiring a staging area or centralized repository (see). Federation is an effective way to capture subsets of very large or very distributed data sets, and is frequently used, when data is offsite, is in an older format, or is infrequently used. For example, a data federation framework will allow an application to access databases, XML data, flat files, or even data services or data streams using a uniform access mechanism. In turn the federation server dynamically accesses the data sources and returns the synchronized results. *Federation simplifies consolidating data from multiple sources, enabling "cross-pollination" of information to better discover opportunities, and is very effective at joining dissimilar data before it arrives to the target.*

- Information Stream processing, which provides applications with continuous access to streaming data sources in real time. *Connecting streaming data with persistent data sources supports complex event processing and real-time discovery of opportunities based on emerging knowledge activities, such as weather-based commodity trading or immediate deliveries to prevent retail out-of-stocks.*
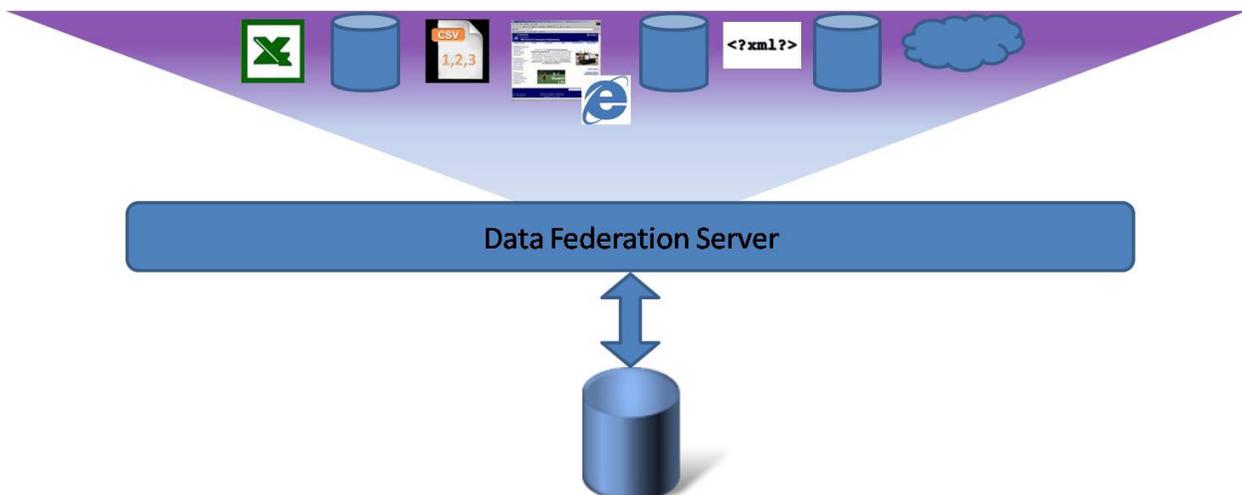
Figure 2: Data federation services provide dynamic, transparent access to numerous data sources.

Cohesive Information Integration

Because most of the data to be analyzed is created upstream from the point of analysis for specific transactional or operational purposes, it is necessary to provide a cohesive information integration framework to facilitate information acquisition and sharing. In other words, preparing enterprise data for analysis requires collaborative techniques for acquiring data from multiple sources, managing representations of similar data concepts in variant representations, and then sharing information with multiple consumers across different applications.

- Data Integration and ETL – Like data federation, data integration relies on the ability to seamlessly access data from many different sources and deliver that data to numerous targets, but it encompasses much more when it comes to parsing extracted data, normalizing it into standard formats, cleansing, and transforming the data into a representation that is suitable for loading into a data warehouse and any subsequent uses for reporting and analysis. *ETL is a key enabler of data warehousing, providing the fundamental information flows that enable business intelligence altogether.*

- Master Data Management – "Master data" objects are those core business concepts represented in different data silos and used in the different business applications across the organization, along with their associated metadata, attributes, definitions, roles, connections, and taxonomies and hierarchies. Some common examples include Customers, Suppliers, Parts, Products, Locations, Agreements, Contact mechanisms. Master Data Management (MDM) incorporates the business applications, information management methods, and data management tools to implement the policies, procedures, infrastructure that support the capture, integration, and subsequent shared use of accurate, timely, consistent and complete master data. MDM should provide the ability to manage access to uniquely identifiable representations for each master data entity across the application infrastructure. *Providing a "unified view" of core data concepts assembled from multiple sources helps to reduce duplication of customers, products, suppliers, or any other key data asset class.*

- Metadata Management – Just the right amount of metadata is necessary in order to support the data extraction, integration, and analysis requirements to ensure consistency of meaning. Metadata incorporates the business definitions (associated with data concepts, business terms, and associated definitions and semantics), conceptual reference data domains and their corresponding values, data element definitions, data element formats, structures, and data types, as well as entity models, entity relationships, and supporting metadata for data governance such as information usage maps, data quality rules, and access controls. *Managing data definitions, reference tables, and data mappings inspires information reuse and helps synchronize the semantics of data as it is integrated from across many systems.*

- **Data Profiling** – Like any asset, it is valuable to inventory the asset – determine what it is, who it belongs to, where it is used, and its serviceability. Data profiling does this – it supports a combination of artifact review and empirical analysis of source data sets to understand the characteristics of data element metadata that are critical for analysis. Data profiling can provide the "ground-truth" associated with the actual data values as well as provide evidence of consistency with metadata, and will provide insight into the suitability of candidate sources to satisfy the target analytical needs. *The result of a data quality assessment using data profiling in conjunction with a review of the consuming application expectations can uncover data quality rules that are managed within the metadata repository.*

- **Data Cleansing** – Data profiling will expose potential anomalies and errors in the data, and erred data used as input to analytical processing will impact the believability of its results. When the source data system owners are able to address design or process flaws that introduce data issues, then the root causes can be eliminated and so can the errors. However, when the source data is outside of the organization's administrative control, processes and technical infrastructure must be in place to parse, standardize, enhance/enrich, and cleanse the data to satisfy the downstream analytic needs. Data cleansing is generally required when the data is being repurposed, especially when the new purposes have stricter data quality expectations. *Data cleansing enhances the value of the data and high quality, consistent data makes the decision-making process trustworthy.*

- **Data Validation** – Alternatively, there are often opportunities for introducing new errors into the data, and identifying and eliminating potential data flaws early in the information production flow will reduce the variance and inconsistency downstream and improve overall operational efficiency. *But even more importantly, validating supplied data with defined data quality rules contributes to the delivery of trustworthy data.*

- **Identity Resolution** – Due to the organic growth of the myriad operational systems deployed across the enterprise, it is not unusual that multiple data instances in different systems represent the same real-world entities in a variety of ways. Alternatively, there are often situations in which two real-world entities share the same identifying attributes, making it difficult to distinguish between them. Both of these types of issues reflect the same core challenge: the ability to evaluate the similarity between pairs of records and determine whether they represent the same thing. Identity resolution addresses these issues by calculating and scoring the degree of similarity between any two records. When the score is above a specific threshold, the two records are presumed to match; below another threshold, they are deemed to not match. Identity resolution is used to match records in the presence of variations or incomplete attributes, or to determine that two records truly represent distinct entities. *Identity resolution is a key component of master data management, and increasing precision in entity matching reduces data duplication and supports high quality reporting and analysis.*

- **Pervasive Delivery Mechanisms** – The spreadsheet is no longer the sole means for delivering analytical results. Self-service configuration of reports and query results from within business

applications eliminates the bottleneck caused by relying on IT staff for support, and web-based delivery simplifies the availability of results using a variety of intuitive, interactive visual presentation objects (such as graphs, heat maps). *By automating event-driven notifications that can be tailored to an assortment of interfaces, ranging from desktops to hand-held PDAs, actionable intelligence can be provided directly where it is needed.*

Analytics Services

The range of analytics services supports the variety of data consumers across the organization:

- Reporting and Ad Hoc Querying – Standard, static reports derived from user specifications provide a consistent view of particular aspects of the business, generated in batch and typically delivered on a scheduled basis through a standard (web) interface. The static nature of reports drives the need for alternative methods for additional insight. One approach is to extract the reported data into spreadsheets for additional data manipulation, while also allowing ad hoc queries to gather additional data for analysis. *Standard reports can provide knowledge to a broad spectrum of consumers, even if those consumers must have contextual knowledge to identify the key indicators and take action. However, given the growth of data into the petabytes, standard reporting is rapidly yielding to exception reporting.*

- Scorecards and Dashboards – If a trained eye is required to scan key performance metrics from canned reports, simplifying the presentation of key performance metrics may better enable the knowledge worker to transition from seeing what has already happened to understanding the changes necessary to improve the business process. Scorecards and dashboards customize an up-to-date presentation of summarized performance metrics, allowing continuous monitoring throughout the day. Pervasive delivery mechanisms can push dashboards to a large variety of channels, ranging from the traditional browser-based format to hand held mobile devices. *Through the interactive nature of the dashboard, the knowledge worker can drill down through the key indicators regarding any emerging opportunities, as well as take action through integrated process-flow and communication engines.*

- Mash-ups – The mash-up takes the dashboard to the next level, allowing the knowledge consumers themselves the ability to identify their own combination of analytics and reports with external data streams, news feeds, social networks, and other web 2.0 resources in a visualization framework that specifically suits their own business needs and objectives. *The mash-up framework provides the "glue" for integrating data streams and business intelligence with interactive business applications.*

- Multidimensional Analysis and Online Analytical Processing (OLAP) – The multidimensional analysis provided by OLAP tools helps analysts "slice and dice" relationships between different variables (within their own hierarchies), such as "what are corporate revenues by time?" or

"What is the availability of products by supplier by location?" The use of the word "by" suggests a pivot around which the data can be viewed, allowing one to look at sales grouped by time period, then by region, or the other way around, grouped by region *then* by time period. *OLAP lets the analyst drill up and down along the hierarchies in the different dimensions to uncover dependent relationships that are hidden within the hierarchies*.

Emerging Advanced Analytics Techniques

In addition to what might be called "mainstream" analytics, there are emerging techniques that are rapidly being incorporated into the environment to support pervasive operational intelligence, such as:

- Stream computing – Streams allow one to analyze every bit of data being sent so that immediate action can be taken when critical information can be inferred. Data can also be selectively saved for subsequent processing. Stream analysis effectively provides a "data federation" capability layered on top of a query engine applied to analyze massive amounts of dynamic data from many heterogeneous sources. *The analyses provided by stream computing enable rapid response to events and changing environments by updating query result sets in lock-step as the data sources are refreshed. In essence, multiple (variant) data sources can be reviewed, filtered, cleansed, aggregated, or subjected to additional transformations to feed numerous downstream data consumer needs.*

- Embedded predictive analytics – Data mining and other advance statistical techniques enable analysts to build models that, to some extent, replicate some of the "human mechanics" and thought processes performed to recognize patterns for success. Analysts using data mining techniques develop predictive models that can be refined, trained, and then applied to very large data sets to identify patterns corresponding to opportunities or risks. *Predictive analytics supplements both operational decision-making and strategic analysis, satisfying the actionable intelligence needs of a variety of knowledge consumers.*

- *Unstructured Data* and text analysis – Text analysis can be used to isolate key words, phrases, and concepts within semi-structured and unstructured text, and these key text artifacts are analyzed semantically, modeled, and their source documents correlated based on recognized concepts. *Once this analysis is completed, the information contained within the documents can be clustered, categorized, and organized to support intelligent searches, and filtering concepts from streaming text helps identify important text artifacts that can be routed directly to individuals with a particular interest in the supplied content.*

- "Real-time" identity recognition – Identity resolution has traditionally been a batch process, but time-critical operations relating to customer service, call center operations, or more sensitive activities involving security, bank secrecy act/anti-money laundering, or other "persons of interest" applications become significantly more effective when individual identities can be recognized in real time. *Real-time identity recognition enables rapid linkage between individuals*

and their related attributes, characteristics, profiles, and transaction histories and can be used in real-time embedded predictive models to enhance operational decision-making.

Challenges

A mature business intelligence and analytics program will have a full complement of these technology components to support knowledge consumers across the full analysis spectrum. In many environments the analytics program has grown organically, with a variety of acquired tools, internally-developed solutions, integrated with different choices of hardware, network, and software (such as database management systems), leading to a workable, if not the most efficient, solution.

And while the horde of vendors have endeavored to support interoperability (so that they all seem to play well together), the BI systems have been engineered over time with heterogeneous components provided by different vendors, with little thought of the complexity of component integration, let alone performance or optimization. In fact, as the need for speed grows, we are recognizing that there are some latent challenges that exist for organizations as they try to "home brew" their comprehensive analytics infrastructure, including these factors:

- **Organic Development and Heterogeneity** – The organic nature of the development means that the analytic applications have been incorporated on an "on-demand" basis, with neither a comprehensive program plan nor an assessment of business needs across the enterprise. This leads to technical dependencies based on development decisions that are not specifically related to addressing business needs, and in time, those dependencies may impede the maturation of a flexible end-to-end analytics solution.

- **Flexibility and Extensibility** – Despite the attempts by many vendors to enable interoperability, each is limited in its ability to work well with those (usually released) product versions for which there are published specifications. In reality, this imposes stringent integration constraints which may prevent the customer from enabling all available product capabilities. For example, if the selected data cleansing tool only works with version 5.7 of the selected ETL product, the customer must refrain from upgrading to version 6 of the ETL product until the data cleansing vendor enhances its product to support the ETL product upgrade. Also, as business needs, requirements, analytics expectations, or numbers of consumers change, the underlying analytics infrastructure will have to adapt to those changes. This suggests a need for an ability to easily add capabilities and functionality to the business intelligence infrastructure.

- **Data quality** – Even with an increased concentration on data governance and quality management, intermittent data validation, different tools for parsing, standardization, and cleansing, and conflicting rule sets *still* contribute to data flaws and inconsistency.

- **Time to value** – Installing, testing, and validating a variety of components and ensuring that they operate well together requires a significant time and resource investment in planning,

design, implementation, and deployment. The increased complexity of implementation and deployment increased the time until the systems can be used productively.

- **Performance and scalability** – Many BI systems become a victim of their own success; as the number of users increases, the query load grows, or as the amount of data to be analyzed grows, the system's ability to scale appropriately leads to performance degradation. This scalability challenge only increases when interoperability constraints artificially throttle back the performance potential of any of the integrated components.

Recommendations for Addressing the Challenges

When considering all these challenges together, a common thread emerges: inefficiencies introduced due to the piecemeal accumulation of technology components from a multitude of sources will ultimately degrade the organization's ability to deliver actionable intelligence to the appropriate individuals at the necessary times, especially as the demand for analytics increases, whether that is due to increased embedded operational intelligence, larger data volumes, increased numbers of users, (more likely) a combination of these demands.

And if a root cause of these risk factors is the variety of technology components, then mitigating those risks requires considering the options for creating an end-to-end solution that is architected to best take advantage of complementary components. A complete suite of tools that support the entire organization's reporting and analytics needs addresses our challenges:

- **End-to-end solution** – With a well-designed architecture, the program team can articulate a strategy for meeting business needs with a comprehensive solution in which all the components are designed to fit together. Selecting a complete solution from a single vendor not only simplifies the implementation and deployment, it also simplifies the procurement process while reducing the risks of a heterogeneous environment.

- **Flexibility and Extensibility** – A single provider can provide greater flexibility, especially when upgrades and releases can be synchronized in a way that ensures that functional improvements are not artificially limited by product versioning. In addition, the customer can introduce functionality as needed by deploying new upgrades or modules in lock-step with the strategic program plan.

- **Data quality** – Standardizing the data validation, cleansing, and enhancement tools and the way those tools are used provides a predictable level of consistency of enterprise data from its initial entry (or creation) to the numerous downstream consumers.

- **Time to value** – Reducing systemic complexity by unifying the complete solution platform simplifies the acquisition process, reduces the resource requirements for implementation, training, and deployment, thereby accelerating the time to value.

- **Performance and scalability** – When an end-to-end solution is designed to run on top of specific hardware, the developers are able to take advantage of a number of optimizations integrated directly into both the hardware and software platforms, such as workload management, task and process scheduling, load balancing, parallel I/O channels, or high availability. Optimized analytical database management services allow for high performance analytical data warehousing, supported by parallelized data integration plus high speed federation services. Increasing numbers of queries can be offloaded to alternate processing units or routed to in-memory databases, decreasing DBMS loads while increasing response rates and throughput.

By transitioning from an organic evolution of corporate business intelligence and analytics environment built on top of a myriad of technology components to a strategically-architected end-to-end solution, your organization can gain a rapid time to value through real-time, integrated analytics, resulting in advantageous intelligence delivered to the appropriate decision-makers at the right time.

About the Author

David Loshin, president of Knowledge Integrity, Inc, (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data quality, master data management, and business intelligence. David is a prolific author regarding BI best practices, via the expert channel at www.b-eye-network.com and numerous books and papers on BI and data quality. His book, "Business Intelligence: The Savvy Manager's Guide" (June 2003) has been hailed as a resource allowing readers to "gain an understanding of business intelligence, business management disciplines, data warehousing, and how all of the pieces work together." His book, "Master Data Management," has been endorsed by data management industry leaders, and his valuable MDM insights can be reviewed at www.mdmbook.com.

David can be reached at loshin@knowledge-integrity.com.