BeyeNETWORK
Global coverage of the business intelligence ecosystem

ANALYTIC ARCHITECTURES:

Approaches to Supporting Analytics Users and Workloads

BI DELIVERY FRAMEWORK 2020

BY WAYNE ECKERSON

Director of Research, Business Applications and Architecture Group, TechTarget, March 2011

SPONSORED BY: VERTICA An HP Company

Executive Summary

THIS REPORT proposes that there are four main types of intelligences designed to turn data into information and information into insights and action: business intelligence (BI), analytic intelligence, continuous intelligence and content intelligence. In the next decade, BI leaders will need to embrace all four types of intelligence to deliver BI applications that deliver lasting value to their organizations.

There is a natural tension between business intelligence and analytic intelligence. The former represents a top-down approach to delivering insights, while the latter represents a bottom-up approach. Business intelligence provides casual users with reports and dashboards populated with metrics that represent business goals and objectives. Analytic intelligence uses ad hoc query tools to evaluate new plans or proposals and answer unanticipated questions. Self-service tools and processes are designed to help close the gap between top-down and bottom-up BI processes but often are not properly implemented, leading to unfortunate side effects.

Continuous intelligence enables organizations to compete on velocity while providing a mechanism to consume ever-larger volumes of data. It also bridges the gulf between data and processes, providing process context to metrics in reports and dashboards and, in some cases, automating processes using analytical rules. Content intelligence provides users with an intuitive interface to explore data of all types, including numeric and document data, without having to first design a schema to house the data.

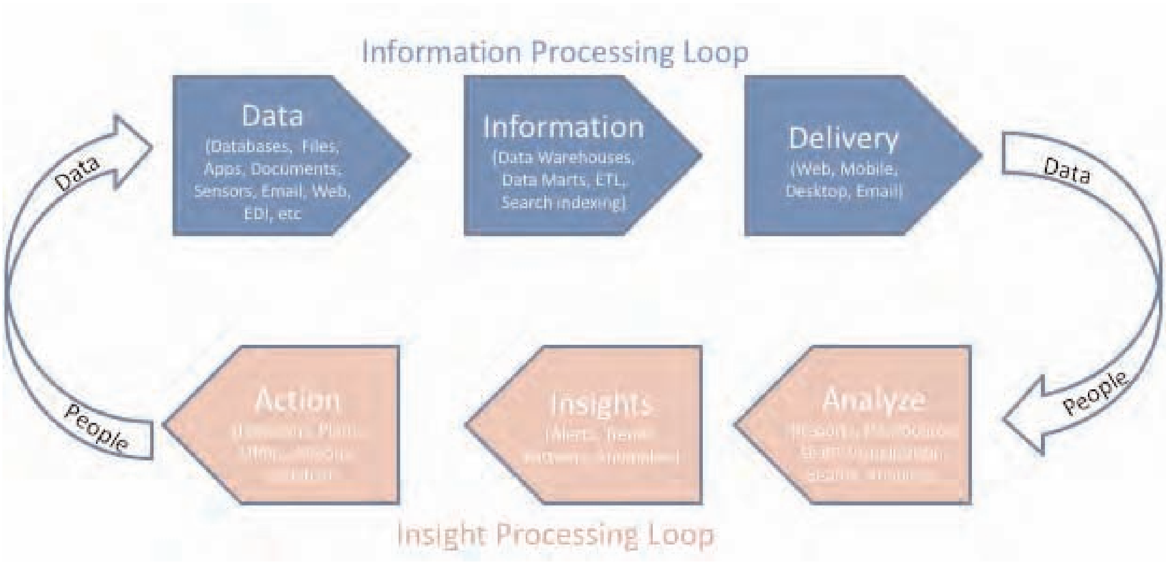
BI Delivery Framework 2020

INFORMATION FACTORY

The BI Delivery Framework 2020 paints a picture of a future business intelligence (BI) environment that converges top-down, metrics-driven dashboards with bottom-up, ad hoc analytics and accommodates both event-driven and unstructured data along with increasingly high volumes of data. The environment pulls data from any source, inside or outside the organization, and delivers data to users via any channel (Web, desktop, mobile or tablet) based on role-based permissions.

The framework fulfills the ideal of the *information factory*, which transforms data into information and information into insights and action (see **FIGURE 1**). This virtuous cycle supports both a learning organization that harnesses information as a competitive advantage and an agile organization that adapts quickly to new events and conditions.

FIGURE 1: ***Information factory***



REPORTING AND ANALYSIS

Strategically, the information factory describes how companies use information to make smarter decisions. Tactically, it is about reporting and analysis, also known as business intelligence, or BI[1]. Organizations hire BI managers, architects, developers and administrators to facilitate the creation of reports and analyses that run against corporate data culled from a variety of internal and external systems.

For more than two decades, BI professionals have tried to shoehorn diverse types of business users, workloads and data types into the same reporting-and-analysis architecture, often with disappointing results. Some users find BI tools too difficult to use, leading to a plethora of BI shelfware. Other users find BI tools too limiting and use them only to populate spreadsheets or desktop databases on which they do their work. In addition, as the velocity of business increases, BI professionals have struggled to deliver timely data through data warehousing architectures designed for batch processing. And these same architectures are now creaking under the load of rapidly rising data volumes and new data types that beg for a continuous approach to data processing. Finally, most BI professionals have yet to figure out how to deliver reports and analyses against 80% of corporate data that isn't found in relational databases, namely documents, Web pages, email messages, social networking data and clickstream data.

To succeed in the next decade, BI professionals need to adopt new thinking and approaches. They need to break away from the "one size fits all" architecture of the past. To meet emerging business demands, they need to manage multiple domains of intelligence and their associated architectures, each of which is optimized for different classes of users and workloads. Without a flexible approach to data architecture, BI professionals will be overrun with requests and the victims of incessant "end-around" plays in which business analysts and departments build their own reporting and analysis environments without the blessing or support of the corporate BI team.

To succeed in the next decade, BI professionals need to adopt new thinking

[1]This report uses the term *business intelligence* as an umbrella term that refers to all the tools, technologies and techniques that support reporting and analysis—and, consequently, make "businesses more intelligent." The term *business intelligence* has been supplanted by other terms, such as *performance management* and, more recently, *analytics*. However, this report still uses *business intelligence* to describe the entire domain and the people who work in it (i.e., *BI professionals*.) The term business intelligence also refers to end-user tools that deliver reporting and analysis functionality. When this report uses the term in that context, it is always followed by the words *tools* or *technologies*, as in *BI tools*.

and approaches. They need to break out of the "one size fits all" architectural approaches of the past.

FOUR DOMAINS OF INTELLIGENCE

This report describes the BI Delivery Framework 2020, which supports the next generation of BI applications (see **Figure 2**, page 7). It defines four domains of intelligence that address unique classes of users and workloads and map to distinct BI architectures, tools and technologies:

1. Business intelligence. This domain addresses the needs of *casual users*—executives, managers, front-line workers, customers, and suppliers. It delivers reports, dashboards and scorecards that are tailored to each user's role and populated with metrics aligned with strategic objectives and goals. This *top-down* driven environment consists of MAD (monitor, analyze, drill to detail) dashboards powered by a classic data warehousing architecture that consolidates enterprise data and enforces information consistency by transforming shared data into a common data model (i.e., schema) and BI semantic layer (i.e., metadata).

2. Analytics intelligence. This domain provides power users—that is, business analysts, analytical modelers and IT professionals—ad hoc access to any data inside or outside the enterprise so they can answer business questions that can't be identified in advance or create various types of analytical models. Traditionally, this type of *bottom-up* analysis has been done in spreadsheets, desktop databases, tools for online analytical processing (OLAP) and data mining workbenches, but is increasingly being conducted in analytical sandboxes running on powerful analytic platforms and databases.

3. Continuous intelligence. This domain automates the collection, monitoring and analysis of large volumes of fast-changing data to support operational processes. It ranges from near real-time delivery of information (i.e., hours to minutes) in a data warehousing environment to complex event processing systems that correlate events emanating from multiple systems and trigger alerts when conditions are met. At massive scale, organizations use specialized event-streaming systems to collect and analyze machine-generated data emanating from sensors and other devices.

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

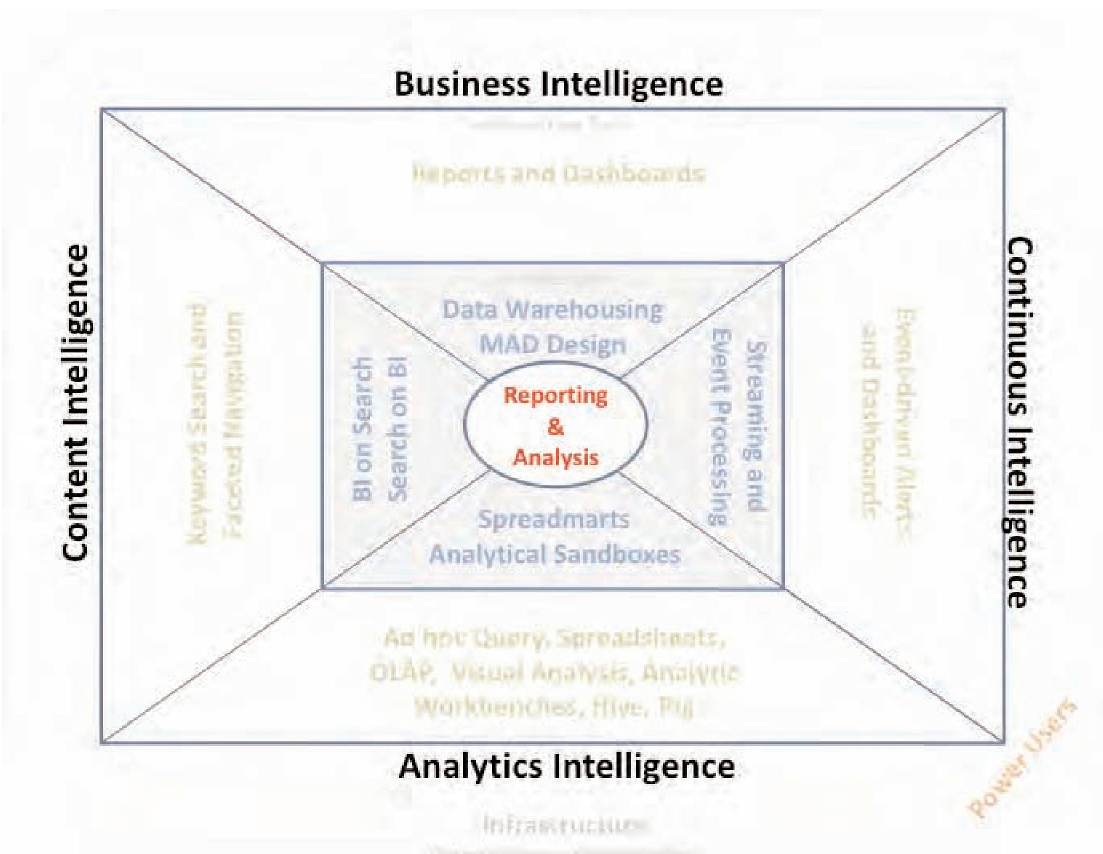
CONTENT INTELLIGENCE

CONCLUSION

4. Content intelligence. This domain gives business users the ability to analyze information contained in documents, Web pages, email messages, social media sites and other unstructured content as well as numeric data found in corporate databases. Content intelligence uses various flavors of search technology (i.e., "search on BI" and "BI on search") to support intuitive access to structured and unstructured data. Specifically, content intelligence uses hybrid search indexes, semantic technology, text mining and faceted navigation, offering significant agility and flexibility in delivering reporting and analysis applications.

The BI Delivery Framework 2020 in **Figure 2** depicts the four intelligence domains and maps them to end-user tools and architectures, described briefly in the bullets above. The four domains are designed to support reporting and

Figure 2: ***BI Delivery Framework 2020***



analysis applications, depicted in the center of the diagram. The framework also shows two overlay dimensions running the length of each diagonal axis. One axis represents the spectrum of analytical styles ranging from Monitor (Top Down) at the top right and Explore (Bottom Up) at the bottom left. The other axis defines the spectrum of business users ranging from Casual Users at the top left to Power Users at the bottom right. The framework also depicts the hardware and administrative infrastructure that encompasses all four domains and is becoming a vital element in the delivery of mission-critical BI applications.

This report will discuss the tools, technologies and architectures employed in each domain. The remainder of this section will drill down on the other dimensions depicted in the framework.

- **Analytical styles.** By mapping analytical styles to intelligence domains, we can see that both business intelligence and continuous intelligence are top-down, monitoring environments that use dashboards and reports to track activity against predefined metrics and targets. In business intelligence, executives and managers track their progress toward achieving strategic objectives and goals using metrics that embody those objectives and goals. In continuous intelligence, operational workers and managers monitor performance against predefined service levels, watching for anomalies and exceptions.

In contrast, analytics intelligence and content intelligence employ a bottom-up, exploratory style of BI. In analytics, business analysts and analytical modelers query, explore and integrate data from various systems and then analyze the information using ad hoc query tools, spreadsheets and analytical workbenches (e.g., SAS or SPSS). In content intelligence, business users also explore the data using keyword search and faceted navigation.

- **Types of users.** The second overlay dimension represents the spectrum of business users. On one end are casual users who use information to do their jobs, while on the other are power users for whom information is their job. Casual users, who consume information produced by power users, prefer a top-down BI environment about 60% to 80% of the time. (For the remaining 20% to 40%, casual users seek ad hoc access to information, which requires self-service BI, and we will discuss that in detail later.)

On one end are casual users who use information to do their jobs, while on the other are power users for whom information *is* their job.

Casual users use the business intelligence domain to monitor performance using business metrics aligned with strategic goals and objectives. They use content intelligence to submit ad hoc queries via keyword search. (Unfortunately, BI search, which is still in its technical infancy, hasn't yet been deployed in a majority of BI environments, robbing casual users of a key intelligence capability.)

In contrast, power users use analytic intelligence to issue ad hoc queries and create analytical models. They use continuous intelligence to manage real-time business activity. (Actually, the analysts devise the rules that correlate events and trigger automated actions, and operational workers monitor the activity in dashboards.)

- **Infrastructure.** Increasingly, BI applications are mission-critical operations, driving processes that run the business. As such, BI applications increasingly require tools and services common in data center operations and high-performance computing that ensure high levels of system reliability, scalability and availability. Such tools and services include virtualization, grid computing, clustering, load balancing, performance and usage monitoring, workload management and scheduling.

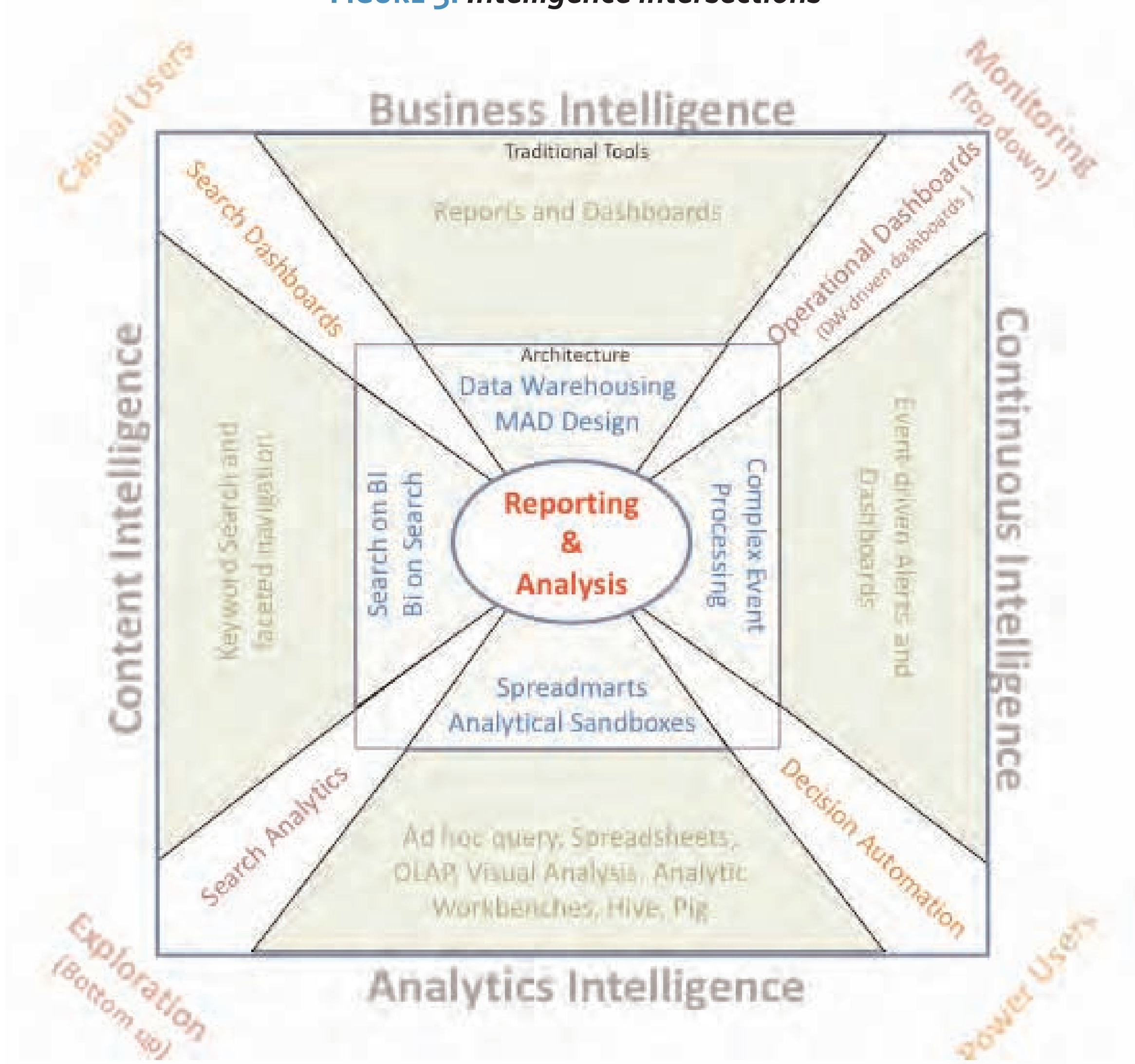
In addition, new advances in computing hardware have caught up with software capabilities and promises. Multicore processors, inexpensive RAM, solid-state disk drives, fast interconnects, storage-based filtering and application processing have dramatically improved BI price and performance. As such, vendors are shipping hardware-software appliances that consist of pre-integrated servers, storage and database software and even BI tools that accelerate time to value and reduce total cost of ownership. These and other technical innovations are revolutionizing the delivery of BI applications, enabling BI teams to deliver existing applications at lower cost with better performance or create analytical applications that previously were cost-prohibitive to build.

INTERSECTIONS

Figure 3 (page 10) shows a slightly different view of the BI Delivery Framework. It depicts BI applications that sit at the intersection between intelligence domains (the diagonal wedges). These applications deliver significant value because they blend multiple intelligences into a single application. (It's also possible to blend intelligences horizontally.)

1. Operational dashboards. Data warehouse-driven, operational dashboards sit at the intersection of business intelligence and continuous intelligence. These dashboards blend the data integrity of business intelligence applications with the timeliness of continuous intelligence applications. The dashboards contain mostly current data with some historical data and are updated at intervals ranging from every 15 minutes to every several hours. Most operational dashboards are built in this fashion. In contrast, business intelligence

Figure 3: ***Intelligence intersections***



dashboards support tactical or strategic applications, not operational ones. They consist of largely historical and summarized data updated daily, weekly or monthly. At the other extreme, continuous intelligence dashboards are event-driven systems that trickle feed data into dashboard objects as events happen, often bypassing the data warehouse altogether.

2. Decision automation. These applications represent a perfect union between continuous intelligence and analytics intelligence. Companies use decision automation engines to drive fast-paced but nonvolatile business processes that have well-known input and output parameters. These systems often embed analytical algorithms to score behavior in real time based on in-the-moment activity, providing real-time, customized responses or offers. In some cases, these engines automate a portion of a process and spit out recommendations for human validation (e.g., fraud detection), or they run independently without human intervention (e.g., Web recommendations, dynamic pricing or personalized gaming).

3. Search analytics. This type of analysis applies the intuitive user interface of search to explore structured and unstructured data in an ad hoc fashion. It combines the ease of use and flexible data access of content intelligence with the ad hoc, exploratory requirements of analytics intelligence. This report investigates search analytics in detail in the next section.

4. Search dashboards. This process blends the intuitive interface and flexible data access of content intelligence with top-down metrics management emblematic of business intelligence. This report explores search dashboards in the next section.

SUMMARY

To succeed with BI in the coming decade, BI professionals must implement, or at least interoperate with, new BI architectures that support new types of users, workloads and data. Embracing the BI Delivery Framework 2020 will enable BI environments to bridge current and future domains of intelligence that are key to delivering on the promise of the information factory. ■

Business Intelligence and Analytics Intelligence

BUSINESS INTELLIGENCE

▪ **Top-down intelligence.** The business intelligence domain delivers reports and dashboards to casual users via a classic data warehousing architecture. It is a top-down driven environment in which business leaders first define the metrics they want to monitor and the questions they want to ask, and then the BI team builds data structures, reports and dashboards to meet these specifications. Specifically, subject matter experts define metrics, dimensions, attributes and navigation paths through the data, and the BI team encodes these into data schema and semantic models built into the data warehouse and BI tools.

The benefit of this top-down approach is that it ensures information consistency—the proverbial "single version of truth"—and avoids disputes over the meaning of common data elements, such as *customer, product* or *sale*. It also embodies the strategy and tactics of an organization in metrics that can be monitored on an ongoing basis, ensuring that individuals, groups and the organization as a whole remain on track to meet overall objectives and goals. A top-down environment is crucial for creating a fact-based decision-making culture that measures performance and holds individuals accountable for outcomes.

▪ **Challenges.** However, it is not easy gaining consensus on rules and definitions for shared metrics and data elements or creating key performance indicators that embody an organization's strategy and goals. For example, business leaders have been known to argue for months about the meaning of the term *customer.* Politics and turf warfare have undermined many business intelligence initiatives at an enterprise level. As a result, the most successful business intelligence applications occur at the departmental level, where consensus is much easier to obtain. Everyone in a department (e.g., finance, marketing or sales) tends to speak the same language and use the same metrics to manage processes. This homogeneity tends to minimize politics and turf warfare.

The other challenge with business intelligence systems is that they often take several months or more to deploy and aren't easy to change. It takes time to define metrics and rules, capture requirements, model schema and semantics, source data and map it to target models, and build reports and dashboards. In an ideal situation, when the BI application is departmental in scope and executives agree on requirements and metrics and data sources contain clean, consistent data, the process may take three months. But in an enterprise deployment, the process can stretch out to nine to 12 months or more. And if the business changes and executives want to ask different questions or use different metrics to track performance, then the BI team needs to cycle through the process again. Business intelligence applications are not ideal to support ad hoc requests and are best deployed in nonvolatile business environments.

The key challenge facing data warehousing professionals today is creating an agile, adaptable data warehousing architecture that keeps up with the business and supports both top-down and bottom-up endeavors. This challenge is especially pertinent in fast-paced business environments and competitive industries where change is constant (see **Spotlight**, "New Ways of Delivering Agile Data Warehouses," page 14).

ANALYTICAL INTELLIGENCE

To support ad hoc requests and analyze unanticipated business issues, businesspeople use a bottom-up approach to reporting and analysis, which is the hallmark of analytical intelligence. By definition, the bottom-up approach is more agile and less expensive than a top-down approach because it doesn't inscribe a business model into the structure of the data.

In a bottom-up world, business users empowered with ad hoc query tools get the data on their own instead of waiting for IT to create a standardized repository of corporate information and associated reports and dashboards. Power users download data into spreadsheets, desktop databases, OLAP databases, visual analysis or analytical modeling workbenches where they clean, standardize, integrate and aggregate data so they can answer pressing business questions. Once the data is in suitable shape, the power users run their analysis and publish the results, usually in a spreadsheet or PowerPoint presentation.

(Continued on page 15)

New Ways of Delivering Agile Data Warehouses

SPOTLIGHT

INTERNET COMPANIES have been particularly innovative when it comes to creating adaptable data warehousing architectures that keep up with changes in the company and support both top-down reporting and bottom-up analysis.

ZYNGA. The online gaming company Zynga splits its data warehouse model into a "standard" set of tables that contain entities that it knows will rarely change, and "nonstandard" entities that can't be anticipated. Standard entities that get hard-coded into Zynga's data warehouse schema include player information, such as name, level, friends and times they were logged in. Nonstandard information includes new products and features that the company's game designers continually dream up. It stores all nonstandard information in a single table as key value pairs. This key-value pair table enables Zynga to continuously adapt its schema to change without having to redesign the schema. Of course, it is not easy to query data stored as key-value pairs. A skilled analyst needs to filter out nonrelevant records in the table and then apply sophisticated query logic to obtain meaningful results.

NETFLIX. Another Silicon Valley pioneer, Netflix, takes a slightly different approach. It believes it can keep its data warehousing schema up to date with changes in the business by eliminating coordination costs, which can cause a five-hour task to balloon into five weeks or more. Rather than trying to coordinate a team of specialists to make the change—from ETL developers and database administrators to data modelers and report developers—Netflix assigns the task to one developer called a spanner, whose knowledge spans all BI disciplines and can make the change rapidly, often in the same day it was requested. Of course, spanners are a rare breed of developer that commands higher salaries than average developers. But they more than pay for themselves in the agility they bring to the BI environment.

NoSQL. Some people say that the best way to create an agile, adaptable reporting and analysis architecture is to abandon the relational paradigm, which depends on a fixed schema that has to be defined up-front. These contrarians generally advocate using NoSQL databases or search indexing technology, which we'll examine in the next section.

The most well-known NoSQL product is Hadoop, a Java software framework that executes tasks in parallel on a distributed file system that runs on a scalable grid of commodity servers. Hadoop stores data in files and is agnostic about how the data is structured. Once developers understand the inherent structure of the data in Hadoop files, they write custom code to query or manipulate that data. Hadoop requires no up-front modeling, but lots of rigorous coding to access data. The reverse is true for relational structures where the heavy lifting is done up-front during the design phase, while data access is streamlined through the use of a common query language, SQL. ■

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

CONTENT INTELLIGENCE

CONCLUSION

(Continued from page 13)

- **Challenges.** By sourcing their own data, business analysts end up performing the same work that IT professionals are hired to do. According to research from The Data Warehousing Institute, business analysts spend on average two days a week managing data instead of analyzing it. They become, in effect, "human data warehouses." This inefficient use of skilled labor costs organizations a significant amount of money, although many executives are unaware of these hidden costs.

Furthermore, in a bottom-up environment, each business analyst creates a unique silo of information. Each analyst uses slightly different rules to define commonly used metrics and data elements. When a top executive calls an operational meeting to discuss results, these analysts often spend hours arguing about whose data is correct, a phenomenon known as "dueling spreadsheets." All hell breaks loose when a CEO asks a simple question, such as "How many customers do we have?" or "What were sales yesterday?" This lack of information consistency at the enterprise level, along with inefficient use of power users, has caused many CEOs to launch data warehousing initiatives as a prerequisite for doing business.

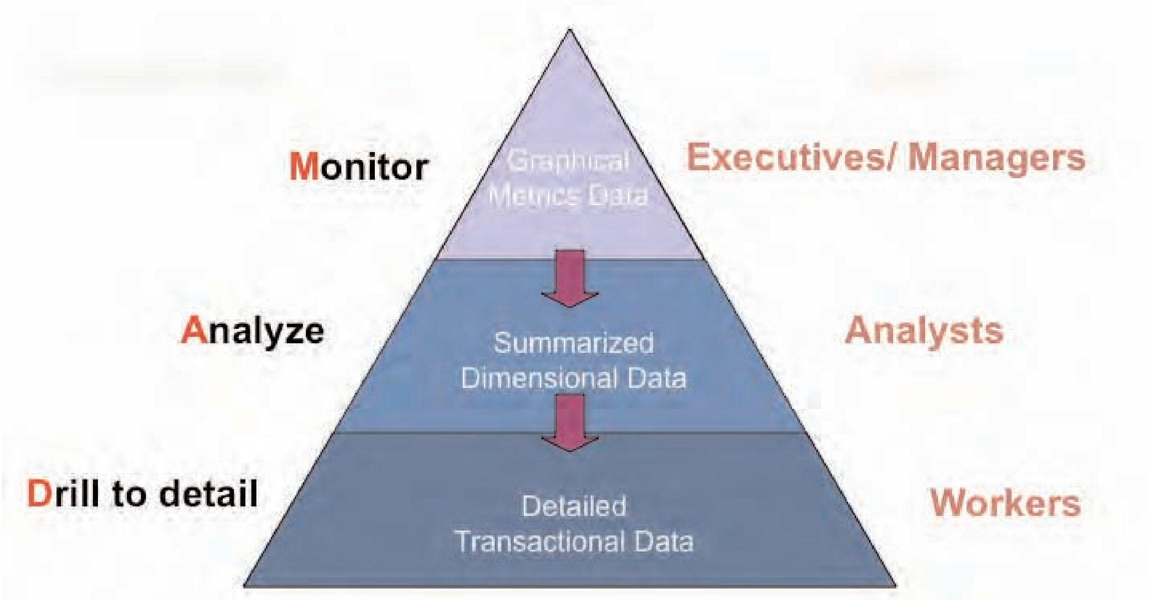
> ***The casual user mantra is 'Give me all the data I want, but only what I need, and only when I need it.'***

BUSINESS INTELLIGENCE ARCHITECTURES

- **Casual-user requirements.** The types of users dictate the style of intelligence and associated architecture. As mentioned earlier, about 80% of the time, casual users want a top-down environment that enables them to monitor key metrics that embody the goals and targets for which they're responsible. They only want to analyze data and drill into detail when there is an exception condition that needs attention. Their mantra is "Give me all the data I want, but only what I need, and only when I need it." In other words, they only want to see high-level data tailored to their role, except when there is a problem. Then they want to see as much relevant, detailed data as possible.

- **MAD dashboards.** Architecturally, casual users want a layered information delivery system that parcels out information on demand. In other words, a

FIGURE 4: ***MAD Framework***



performance dashboard backed by an enterprise data warehouse[2].

The optimal way to design a performance dashboard is to use the MAD framework (see Figure 4). MAD stands for monitor, analyze and drill to detail. These three sets of functionality correspond to three levels of data: graphical metrics data (i.e., stoplights), summarized dimensional data (i.e., analysis tools with dimensional navigation or dynamic filtering), and detail detailed (e.g., operational queries or reports). Users can enter at any level of the framework and navigate upward or downward. Executives and managers tend to spend more time at the top level, analysts at the middle level and front-line workers at the bottom level.

- **Role-based views.** A well MAD dashboard tailors the metrics and views at each level to each user's role and tasks. It only shows them what they need to see and discards the rest. These role-based views can be delivered as different tabs within a single dashboard or a series of linked dashboards. In either case,

[2]See Wayne Eckerson, "Performance Dashboards: Measuring, Monitoring, and Managing Your Business," Wiley & Sons, Second Edition, 2010.

SUMMARY
BI DELIVERY FRAMEWORK 2020
BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE
TOP-DOWN VERSUS BOTTOM-UP
CONTINUOUS INTELLIGENCE
CONTENT INTELLIGENCE
CONCLUSION

the goal of a MAD dashboard is to design once and deploy multiple times using role-based views. Most dashboard development platforms support this design paradigm.

- **Metrics.** The shape of the pyramid represents the number of metrics and amount of data at each level. In most MAD dashboards, there are about 10 metrics at the top level. These 10 metrics are filtered by about 10 dimensions at the next level, creating about 100 metrics. And these 100 metrics are each filtered by another 10 dimensions at the bottom level, creating 1,000 metrics all together. This creates a suitably sized sandbox for casual users: It's not so big that they'll get lost and not so small that they'll hit the boundaries too quickly.

- **Visualization.** It also leverages visualization techniques to communicate the meaning of data quickly. At the top level, users quickly glance at graphical metrics to understand the status, trend and variance of key performance indicators. If something is awry, users click on the metric and drill to the next level to perform root cause analysis. If they need to know which customers or products are affected by a problem, they can drill to the lowest level of detail available, such as orders and order items in an operational sales report.

A MAD dashboard creates a suitably sized sandbox for casual users: It's not so big that they'll get lost and not so small that they'll hit the boundaries too quickly.

Although MAD dashboards represent a classic top-down application in the business intelligence domain, many departments build MAD-like dashboards using bottom-up visual analysis and search tools. These "bridge" tools deliver graphical metrics at the top level and an intuitive and flexible analysis and discovery at the middle layer. Some visual analysis tools even incorporate advanced analytics, enabling users to apply regressions or other algorithms to forecast outcomes and cluster or classify customers, among other things. However, both visual analysis and search tools often have trouble delivering the bottom layer of detailed data in a report format that users are accustomed to seeing. This makes sense since they are analysis and discovery tools, not reporting tools.

ARCHITECTURES FOR POWER USERS

- **Liberation and proliferation.** Power users have very different architectural requirements than casual users. While casual users want and need information boundaries, true power users balk at having any boundaries at all. They want the freedom to grab any data they need and manipulate it however they want. They are usually under intense pressure to answer a business question from a top executive and can't afford to wait for the BI team to deliver a sanitized data set in a top-down fashion. So they create their own data silos, causing a proliferation of spreadmarts and renegade data marts that undermine corporate information consistency.

- **Performance hogs.** At the same time, power users are big users of data warehouses, almost too big. They issue complex, long-running queries that often bog down system performance of the data warehouse for casual users who are simply trying to view dashboards and reports. The tactical fix many BI leaders apply to this problem is to pre-run reports for casual users at night or restrict power users' ability to submit complex queries until after hours. Neither application is ideal.

> ***Power users are usually under intense pressure to answer a business question from a top executive and can't afford to wait for the BI team to deliver a sanitized data set in a top-down fashion.***

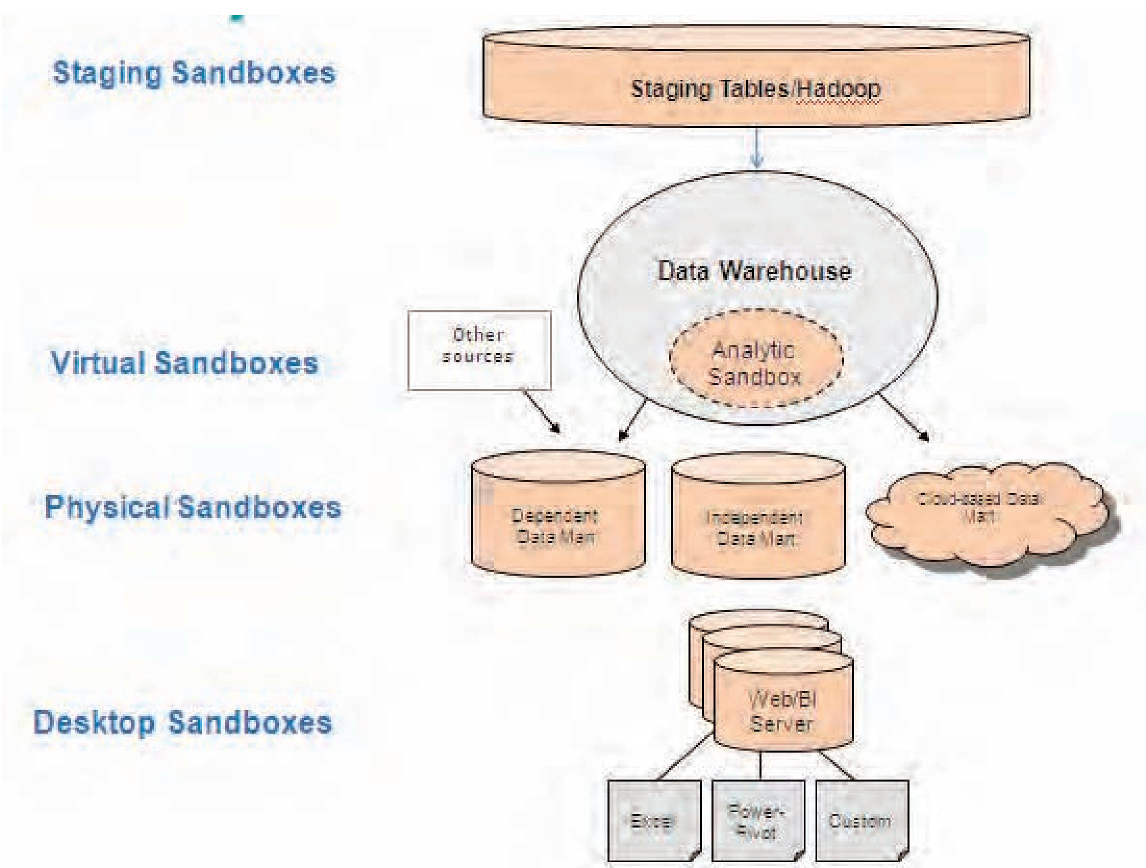
- **Analytical sandboxes.** The ideal architecture for analytic intelligence gives power users the flexibility to mix and match any data they want without having to create spreadmarts that undermine information consistency. Companies are doing this by implementing one or more analytical sandboxes (not to be confused with the MAD sandbox mentioned earlier). An analytical sandbox gives power users (mainly business analysts and analytical modelers) a safe zone within the data warehousing environment—or as one BI director called it "a playground"—to merge, explore and analyze data to their hearts' content, without interfering with performance of other workloads on the system. The only thing analysts can't do in these analytic sandboxes is publish data (i.e., reports and dashboards) for general consumption.

There are many types of analytic sandboxes (see **Figure 5**).

1. Staging sandboxes. With staging sandboxes, power users are given access to a landing area for source data before it is standardized, integrated, summarized and loaded into the data warehouse. Only the most skilled and trusted business analysts access staging areas because the data is still in its raw form. Staging areas store data in a relational database or file system.

The advantage of using a staging area is that it brings almost all of the raw, detailed data that an analyst might want to mine in a single place. This saves the time and hassle of having to access each source separately and merge the data locally. A disadvantage of a staging area sandbox is that analysts usually

Figure 5: ***Types of analytic sandboxes***



aren't allowed to upload data to the staging area, which is a problem if the staging area doesn't contain all the data sources the analyst requires. Also, since the data is in its raw, nonintegrated form, it takes an analyst with extremely strong SQL skills and knowledge of source system schemas to understand and manipulate the data properly. Finally, analyst queries may interfere with data transformation or aggregation processes running against the data.

- **Hadoop.** Many companies now use Hadoop, an open source distributed file system, as a data warehousing staging area and a general purpose analytical sandbox. Today, many companies use Hadoop to stage, transform and summarize clickstream and other large volumes of nonstandard data, such as Twitter feeds and sensor data, before loading the data into the data warehouse for analysis. Many also permit analysts to run standard reports or ad hoc queries directly against Hadoop using custom-developed Java programs because it would be cost-prohibitive to load all the detail data into the data warehouse and perform the analysis there.

However, as a new open source software, Hadoop lacks data center services common to most enterprise computing environments. Companies looking to implement Hadoop should work with vendors, such as Platform Computing, that provide cluster, grid and cloud management tools for Hadoop and other scale-out compute infrastructures.

2. Virtual sandboxes. These environments are partitions within data warehousing databases created by database administrators (DBAs). Business analysts can upload their own data into the partitions and mix it with data from the data warehouse for exploration, analysis, testing and prototyping. DBAs either allow the analyst to query select data warehousing tables or push selected data into the virtual partition via an ETL process. DBAs can create partitions for individual analysts or a single, larger partition for all analysts if they want to share their work. DBAs may need to apply workload management controls to prevent analyst queries from affecting performance of other applications and workloads running in the data warehouse.

The primary advantage of a virtual sandbox is that it brings analyst activity out into the open under the watchful eye of the IT department, which can monitor activity. Another advantage is that it doesn't duplicate data warehousing data on other servers, preventing corporate information from getting out of sync. It also eliminates the need for a duplicate system and the associated hardware, software and operating costs required to maintain a mirror copy of the data. However, virtual sandboxes may require administrative finesse to prevent analyst queries from bogging down the performance of other workloads running on the data warehouse, especially if the virtual sandboxes prove popular with analysts. Defining processing priorities and allocating appropriate resources to each workload can be tricky and may require consulting assistance.

3. Physical sandboxes. These environments are physically detached data marts or data warehouse replicas created to offload complex queries from the production data warehouse or to house new types or sources of data that can't be stored in the data warehouse because there is no room. Many physical sandboxes are built on analytic platforms, which are designed to speed the processing of simple and complex queries against large volumes of data. These new platforms, which often come in the form of all-in-one appliances, turbocharge query performance by leveraging massively parallel processing, columnar storage and compression, storage-level filtering and in-memory caches. When used in combination, visual tools and analytic databases provide a compelling bottom-up analytic application that is quickly becoming the industry norm.

Business analysts can upload their own data into the virtual sandbox partitions and mix it with data from the data warehouse for exploration, analysis, testing and prototyping.

Although physical sandboxes replicate data and increase operating costs, as mentioned above, they often are a suitable option in high-volume data processing environments. For example, the online gaming company, Zynga Inc., which produces "Farmville," "Mafia Wars" and other popular Facebook games played by tens of millions of people each day, streams tens of billions of records daily into two mirror copies of its data warehouse, each running on 100-plus node Vertica clusters. One cluster supports production reporting in which access is tightly controlled, and the other supports ad hoc queries from highly trained analysts. All other employees query a subset of the analytical data warehouse that comprises about 1%

of the total volume of data. By using mirror clusters, Zynga's data warehousing environment supports both top-down and bottom-up processing.

Some companies build physical sandboxes in the public cloud. These cloud-based sandboxes are ideal for development, testing, prototyping and analysis on subsets of data. Users can provision their own servers and upload data while DBAs can populate the sandboxes with copies of data warehousing data. Cloud-based sandboxes eliminate the costs and delays associated with purchasing, installing and maintaining additional servers. The pay-as-you-go model avoids the need to justify new capital expenditures.

4. Desktop sandboxes. The most controversial type of sandbox is the desktop sandbox. Ostensibly, this is no different than a spreadmart, except that new BI tools have better controls to govern the dissemination of information and maintain audit trails of all activity. For example, Microsoft PowerPivot is a spreadsheet on steroids that encourages developers to publish data to Microsoft Sharepoint, through which authors and Sharepoint administrators can govern who can access the data and how. Visual analysis tools, such as Tableau Software and QlikTech's QlikView, are in-memory desktop tools that enable business analysts to source data from any system, visually explore data at the speed of thought and publish visually appealing views to casual users via a shared server-based environment. And for more than a decade, power users have created OLAP cubes to examine data dimensionally.

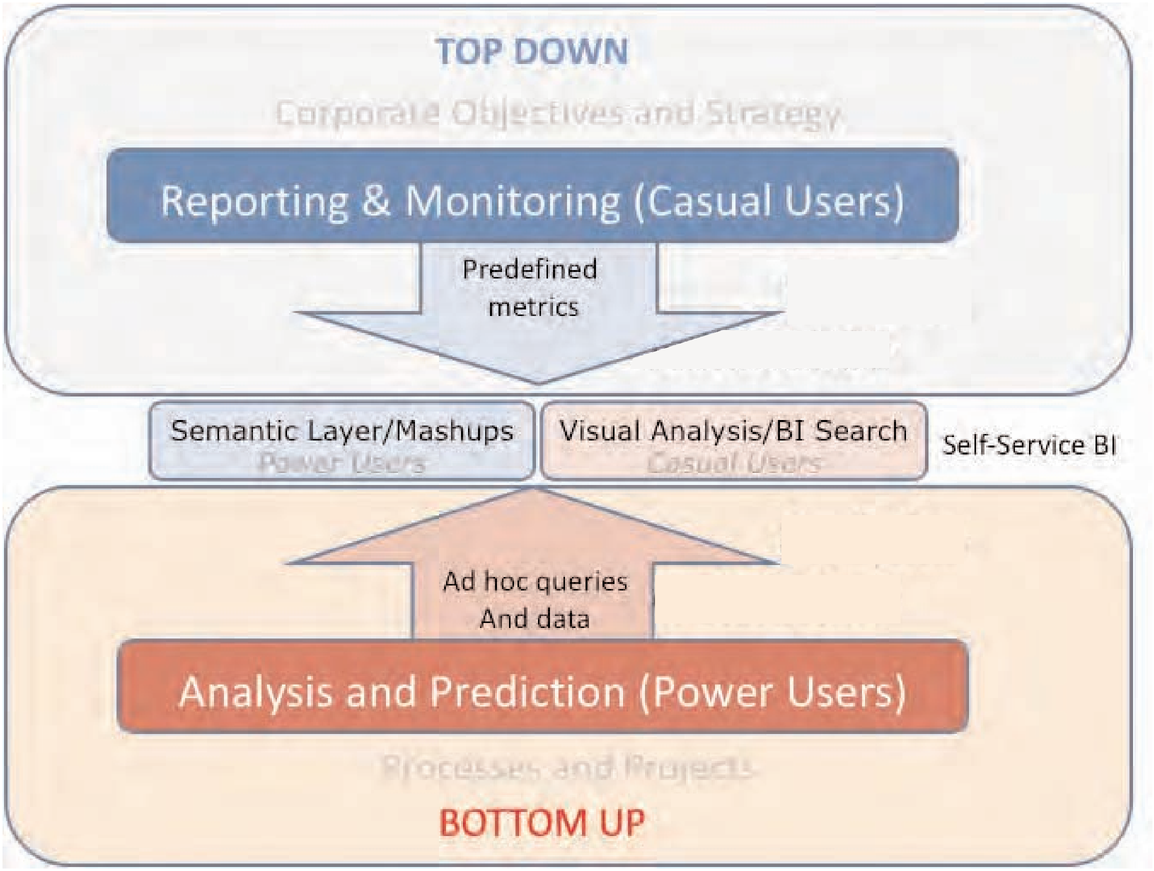
The benefit of a desktop sandbox is that it gives the analyst full control over the sourcing and manipulation of the data but maintains an audit trail of everything they do and gives casual users access to the results of their analysis in a shared, controlled environment. If administrators see that multiple analysts are pulling the same data sets or creating the same reports, they can encourage analysts to leverage one another's work or, better yet, add the data sources and reports to the standard data warehousing environment. On the downside, these sandboxes don't always prevent the proliferation of nonstandard data sets. ■

Top-down Versus Bottom-up

FIGURE 6 shows the dynamic between top-down and bottom-up approaches to reporting and analysis. The mistake that organizations make is to apply either a top-down or bottom-up approach to all reporting and analysis tasks. In reality, organizations need to implement both approaches and apply appropriate governance to ensure that one doesn't bleed into the other.

A top-down approach enables casual users to consume reports and dash-

FIGURE 6: ***Top-down versus bottom-up intelligence***



boards aligned with strategic objectives and goals. The bottom-up approach enables power users to issue ad hoc queries to analyze issues and trends associated with business processes and projects. In short, the top-down approach delivers reports, and the bottom-up approach delivers analysis.

Reports compile information about the way things are done, while analysis gathers information about things that are new or changing. In essence, analysis focuses on change, while reports focus on order. This dynamic is as old as human history. Humans and organizations that support a fluid interplay between order and change are able to adapt and grow gracefully, while those that don't collapse under their own weight, usually a victim of too much rigidity (i.e., order, tradition, law) or too much chaos (i.e., change, freedom, individuality). In the world of BI, organizations need to balance reporting and analysis and not tip too far in either direction.

The challenge in BI is to build a bridge between the disparate worlds of top-down and bottom-up BI. Typically, this is done manually. Power users analyze data to understand what's important to measure in reports. They then create reports for casual users based on the analyses they have done. Thus, analysis begets reports in a two-step process. Power users become the "producers" and casual users become the "consumers."

Unfortunately, turning power users into full-time report developers is not a good use of their time. They get hired to generate insights, evaluate proposals and create models, not deliver reports. Although BI professionals get paid to develop top-down, production-oriented reports and dashboards, they often get overwhelmed with requests for custom ad hoc reports. To bridge the gap, savvy BI teams often recruit super users—a type of power user—to meet the need for ad hoc reports and views.

- **Super users.** Super users are technically savvy businesspeople in each department who become proficient in the use of BI tools. They quickly become the go-to people in each department, creating custom reports on behalf of their casual user colleagues. Since they are embedded in the business and know the processes and people intimately, they become efficient and effective extensions of the BI team. Besides offloading ad hoc reporting duties, volunteer super users become the eyes and ears of a corporate BI team in each department and facilitate the efficient and effective delivery of comprehensive BI applications. BI leaders who want to manage a successful BI program need to identify these super users and work closely with them.

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

CONTENT INTELLIGENCE

CONCLUSION

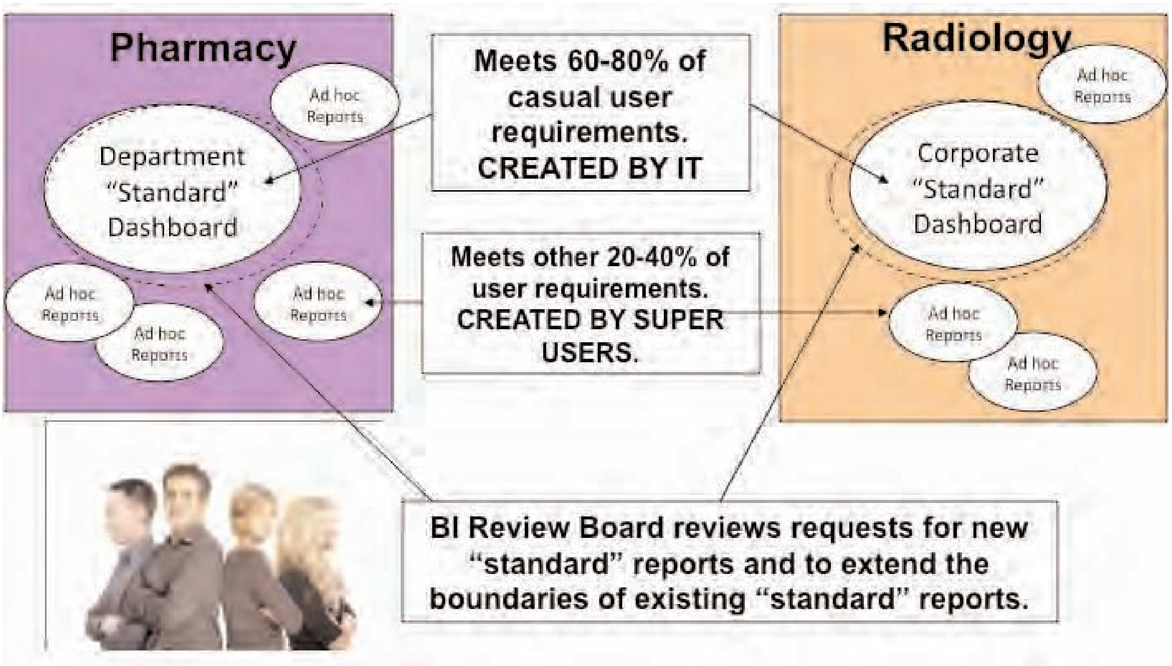
■ **Report governance.** Super users are key to effective BI governance and managing a healthy balance between top-down and bottom-up approaches.

Ideally, the corporate BI team in conjunction with departmental super users creates a set of "standard" top-down, BI reports and dashboards. If designed correctly, these standard reports should meet about 60% of the information needs of casual users (but not power users.) The remaining 40% of requirements are impossible to anticipate—they are bottom-up, ad hoc inquiries. But since casual users by definition aren't capable of generating their own reports and dashboards, they turn to super users to meet their needs.

Besides fulfilling requests for ad hoc reports, super users should also review requests for new standard or official reports. Corporate BI teams need to appoint super users from each department to serve on a report review board that maintains an inventory of existing reports, identifies overlaps between new and existing reports, and makes recommendations whether a new report should be built or an existing one expanded (see FIGURE 7).

Putting super users in charge of both ad hoc and conventional reporting is an effective way to prevent report chaos and the lack of information consis-

FIGURE 7: *Report governance*



tency that accompanies it. Many people equate this strategy as akin to putting the "fox in charge of the henhouse" since super users (or power users in general) are the major culprits behind the creation of renegade BI systems. Yet, BI teams that have adopted this strategy find that distributing or "giving up" control is a powerful way to create key allies and expand BI's footprint in the organization.

SELF-SERVICE BI TOOLS

For the past decade, a robust super-user network has been the only way to make the concept of self-service BI a reality. In theory, self-service BI empowers business users to create their own reports and conduct their own analyses without IT or power-user involvement. In reality, most self-service BI tools have been too hard for casual users to use and too easy for power users to abuse. Paradoxically, self-service BI has often led to large volumes of BI tool shelfware on one hand and report chaos on the other. BI professionals who embrace self-service BI without understanding its potential consequences are really practicing "self-serving" BI—they are eager to offload custom report creation duties at any cost.

> ***Most self-service BI tools have been too hard for casual users to use and too easy for power users to abuse.***

Between the two top-down and bottom-up approaches depicted in FIGURE 6 are BI tools designed to support self-service BI, which has been the elusive Holy Grail of BI for many years. Some tools emanate from the top-downside (e.g., semantic layers and mashups), while others hail from the bottom-up side (visual analysis and BI search).

- **Two types of self-service.** One problem with implementing self-service BI is that BI professionals don't recognize that there are two types of self-service BI, one for casual users and another for power users. FIGURE 8 (page 27) shows two hierarchies of self-service capabilities based the type of user. Users traverse the hierarchy from top to bottom as they gain more experience and confidence with the tools. For example, casual users increase their ability to interact with and analyze data as they traverse the hierarchy, while power

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

CONTENT INTELLIGENCE

CONCLUSION

FIGURE 8: ***Types of self-service BI***

users learn to create more sophisticated reports and views.

Executives, for example, may start out simply viewing static reports but over time progress to navigating predefined drill paths. Managers may start by navigating data and learn to modify existing data sets (sort, rank, filter, add or delete columns), and possibly explore new dimensions of data and create what-if models. Power users, on the other hand, may start by personalizing views for colleagues and assembling reports and dashboards from pre-existing report parts (i.e., mashups). They then may craft new reports using metadata (i.e., semantic layer) and possibly source data independently and develop new applications using a scripting language.

The key with self-service BI is to expose these capabilities on demand as users are ready to use them. On one hand, it's important not to overwhelm users with unneeded and unwanted capabilities; on the other, it's important not to frustrate users who seek functionality that isn't available. The good news is that many BI vendors now recognize this dynamic and do a reasonable job of exposing self-service functionality on demand.

Ironically, top-down BI delivers self-service functionality geared to power users, while bottom-up BI delivers self-service functionality geared to casual users. Let's explore these types of self-service BI.

TOP-DOWN, SELF-SERVICE BI

- **Semantic layers.** Most BI tools require architects to build a semantic layer that models data artifacts in the data warehouse (or any other source for that matter) in business-friendly terms. In principle, a business user armed with a semantic layer and a wizard-based, query-generation tool should be able to construct well-formed queries against the data warehouse and generate their own reports without IT assistance. Unfortunately, most casual users find this process too difficult or time-consuming. This frustrates BI managers who believe self-service BI can alleviate their backlog of custom reports. The good news is that while casual users won't use a semantic layer, super users will.

- **Mashups.** In the past few years, many BI vendors have deployed mashups or mashboards that enable super users to create custom dashboards for themselves or colleagues by dragging and dropping prebuilt widgets from a library onto a dashboard canvas. The widgets are predefined report parts—such as tables, charts, or filter—built using the vendor's report design tool. The widgets share a common interface, akin to Google Gadgets, which enables the widgets to interoperate in a dashboard environment. For instance, if two widgets or charts contain data that share a common key, they will stay synchronized. So, if a user filters the view on one widget, the other automatically updates to reflect the change. Many mashup tools also allow super users to connect to external Web pages using URLs.

> ***A semantic layer creates 'guardrails' for accessing data; this simplifies access but creates problems if users want to go 'off-road.'***

The problem with semantic layers and mashups stems from their top-down orientation. Semantic layers require you to know up-front what data users want to query and how they want to query it. In essence, a semantic layer creates "guardrails" for accessing data; this simplifies access but creates problems if users want to go "off-road." With mashups, professional report designers first need to create a report and then widgetize its components and place them in a library. Both semantic layers and mashups assume that BI managers know what users want to see before they see it.

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

CONTENT INTELLIGENCE

CONCLUSION

BOTTOM-UP APPROACHES TO SELF-SERVICE BI

Bottom-up approaches to self-service BI avoid some of the traps of top-down approaches. Visual analysis and BI search tools impose fewer constraints on what data users can see or how they navigate through the data. Visual analysis tools are visualization tools that enable users to sift through in-memory data at the speed of thought, while BI search tools provide flexible navigation within any data indexed by a search engine. With both toolsets, there are no guardrails inside the data sets that may restrict what users see or how they see it.

Visual analysis tools enable users to apply filters dynamically so they update all objects on the screen instantaneously, making it easy to see correlations and navigate to new and unanticipated views of the data. Similarly, BI search tools enable users to navigate across multiple data sets, both structured and unstructured, using keyword and faceted search to filter and refine their views as they go along.

Bottom-up approaches falter at the enterprise level, where there is an imperative to maintain common definitions for shared data elements.

The tools also align with the "analysis begets reports" dynamic mentioned earlier. Once power users navigate to a view that is particularly illuminating, they save the view and publish it to others. If the issue is an ongoing concern, they may schedule the view to refresh on a regular basis. They may even modify the view to make it more easily consumable by casual users by hiding certain fields or functionality, or by redesigning the display to mimic the look and feel of other reports or dashboards in the company.

- **Constraints.** Bottom-up approaches falter at the enterprise level, where there is an imperative to maintain common definitions for shared data elements. Most visual analysis and search tools are currently deployed to support departmental initiatives or one-off applications when there is no pressing need to establish consensus among shared data elements across the enterprise. While there is no reason these tools can't be used to support enterprise deployments, they don't have the built-in architecture to enforce information consistency. They would rely on a data warehouse to do the heavy lifting for them.

SUMMARY
BI DELIVERY FRAMEWORK 2020
BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE
TOP-DOWN VERSUS BOTTOM-UP
CONTINUOUS INTELLIGENCE
CONTENT INTELLIGENCE
CONCLUSION

Also, in-memory visual analysis tools are constrained by the amount of data they can hold in memory. Despite the advent of 64-bit operating systems and cheap RAM, this constraint may preclude using these tools for large-scale, enterprise deployments, at least today.

- **BI search—ad hoc for casual users.** The ultimate self-service BI enables casual users to ask any question of any data without IT or power-user intervention. To date, this type of environment has not existed. However, new search technology promises to make such self-service BI a reality. As we'll see in a later section, BI search tools enable users to navigate across multiple data sets, both structured and unstructured. They simply type queries in plain English into a keyword search box and then refine the result sets by clicking on dynamically generated categories, called facets. This intuitive interface made popular and familiar by Google, Yahoo and other search engines may finally empower casual users to fully service their own information requests.

SUMMARY

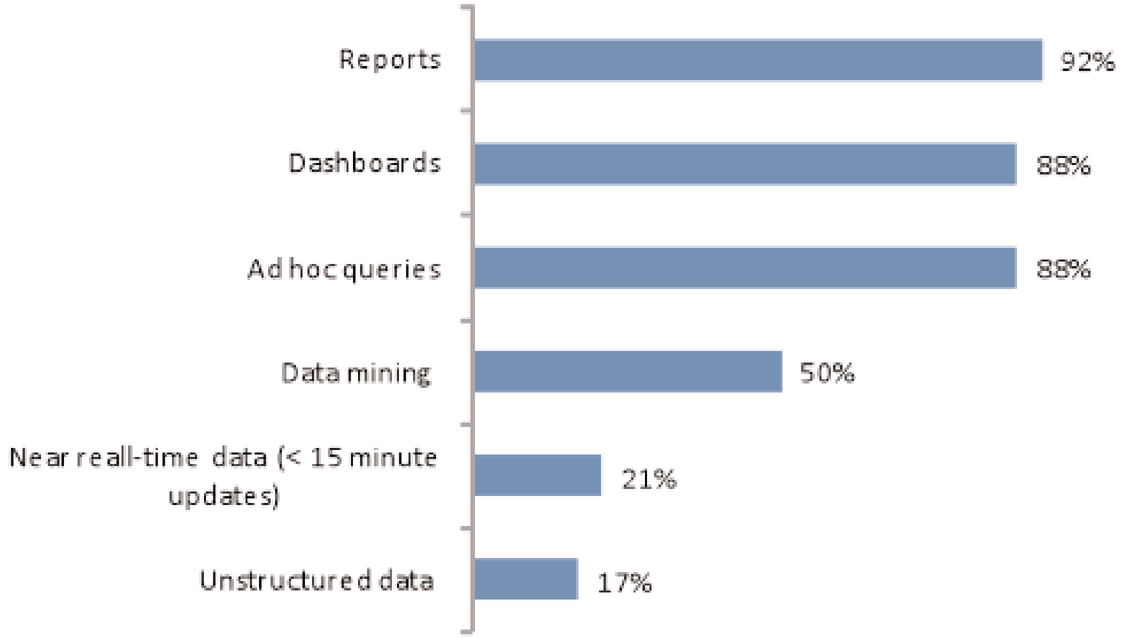
Most companies make the mistake of trying to shoehorn all BI activities into a top-down or bottom-up BI environment. In reality, BI teams must balance both top-down and bottom-up approaches to BI. They must recognize the value that each approach offers to business users and deploy the right tools and architectures for the right users. They also must create a robust super user network and deploy self-service BI tools that expose functionality on demand. They must also rearchitect their data warehouses and change management processes to adapt more quickly to new business requirements so the data warehouse can support both top-down and bottom-up approaches to BI.

Continuous Intelligence

TRADITIONALLY, DATA warehouses are refreshed with current data at night or on the weekend. This rate of replenishment is woefully inadequate for people or departments that manage operational processes and need to stay abreast of events as they happen. As a result, data warehouses have failed to provide much operational support.

However, many executives now recognize the importance of empowering their staff with continuous intelligence. Today, about 21% of organizations update their data warehouses every 15 minutes or less, a signature of a continuous intelligence environment (see FIGURE 9). This percentage has climbed

FIGURE 9: ***Survey question: "Our data warehouse handles a majority (75%+) of the following workloads:"***



SOURCE: BASED ON A SURVEY OF THE BI LEADERSHIP FORUM, AN ONLINE GROUP OF BI DIRECTORS, JANUARY, 2011. WWW.BILEADER.COM/BI_LEADERSHIP_FORUM.HTML.

from 5% in 2007[3] and will likely continue to grow until continuous intelligence becomes a mainstay of BI environments.

- **Strategic value.** One reason BI leaders should adopt continuous intelligence is strategic: Companies today compete on velocity. The difference between success and failure in many industries depends on how quickly companies react to events. Several years ago it would take a day to update a supply chain; now it takes 10 minutes. Call centers used to turn around inquiries in eight hours; now they do it 10 seconds; airlines used to track departures and arrivals every 20 minutes; now they do it every 30 seconds; Wall Street traders who could execute trades in less than 20 milliseconds were kingpins; now the bar is less than one millisecond.[4] And so on.

For example, 1-800 Contacts, an online provider of contact lenses, several years ago created an operational dashboard updated every 15 minutes to help call center managers and salespeople monitor sales and orders against goals. The operational dashboard replaced a set of daily reports that had little impact on performance because the data was not current or accurate enough. The new near-real-time dashboards turbocharged call center productivity, generating a significant uplift in sales and providing the company with a competitive advantage in the industry.

- **Tactical value.** The other reason is tactical. The volume of data that business people want to analyze is growing larger than the batch window available to process and load it into the data warehouse. As a result, the only way to keep up with the desire for more detailed transaction data and new types of data, including clickstream, sensor and sentiment data, is to process the data in near real time. In other words, large data volumes require near-real-time processing.

- **Big data.** The era of "big data" is upon us. Business systems today capture minute details of customer activity and business operations. Rather than analyze call detail records in aggregate, analysts at telecommunications companies want to compare traffic and usage patterns by pairs of originating and

---

[3]See the report I wrote in 2007 titled "Best Practices in Operational BI: Converging Analytical and Operational Processes," which can be downloaded from The Data Warehousing Institute at http://tdwi.org/research/list/tdwi-best-practices-reports.aspx.
[4]Source: Roy Schulte, Gartner Inc.

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

CONTENT INTELLIGENCE

CONCLUSION

destination phone numbers to optimize pricing. Companies now want to track customer sentiment by scraping data from countless Web forums, blogs, websites and Twitter feeds. They want to monitor the status of orders and shipments by tracking products flowing through their supply chains via radio-frequency identification readers. Insurance companies and trucking firms want to evaluate the performance of drivers and their vehicles by capturing a continuous stream of data beamed from dozens of on-board sensors.

- **Process intelligence.** Perhaps more important, continuous intelligence promises to reunite data and process, which have steered divergent paths for the past 20 years. Analytical systems essentially strip the process context from data. Metrics stored in data warehouses and displayed in reports record the output of a process but don't reveal anything about the nature of the process and its anomalies. When a metric trends downward, users may drill into a report or dashboard to find the root cause among other metrics. While BI tools may identify the vicinity of the problem, they often don't expose its source. That requires phone calls and detective work to find out why a process broke down. Maybe it was because Suzie was on vacation and didn't adequately train her replacement on how to handle a flood of last-minute orders requiring custom shipment.

Continuous intelligence promises to close the gap between data and process in numerous ways. One, it exposes performance outcomes in near real time so problems are more quickly detected and investigated. Second, it instruments processes to a finer degree, monitoring each step in a process and correlating results using rules defined from past outcomes and behavior. In some cases, it can automate decisions and actions by embedding rules and analytics into customer-facing applications.

Given the big data and process imperatives for delivering near real-time data, BI leaders need to take steps now to deliver continuous intelligence. The good news is that that there are many ways to implement continuous intelligence. The options generally fall into two camps based on where the bulk of the processing is done: inside the data warehouse or outside of it.

INSIDE THE DATA WAREHOUSE

The benefit of performing continuous intelligence inside the data warehouse is that it enables the organization to maintain a single, clean, consistent set of

SUMMARY
BI DELIVERY FRAMEWORK 2020
BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE
TOP-DOWN VERSUS BOTTOM-UP
CONTINUOUS INTELLIGENCE
CONTENT INTELLIGENCE
CONCLUSION

enterprise data for all reporting and analysis tasks. The challenge here is turning a batch-processing system into one that is dynamically updated.

- **Mini-batch loads.** The first step that most BI teams take in this journey is to accelerate the batch-loading cycle from daily to hourly to every 15 minutes. Instead of doing a single batch load at night or on the weekend, BI teams perform more batch loads with smaller volumes of data, often using change data capture mechanisms to update only those data elements that have changed since the prior interval, instead of refreshing all data from scratch. Depending on the volume of data, BI administrators can drive these mini-batch loads down to a few minutes.

- **Event-driven trickle feeds.** At some point, however, mini-batch loads may not be fast enough to keep up with expanding data volumes. In this case, BI leaders need to invest in an event-driven system that trickles data into the data warehouse as events occur. There are many ways to architect event-driven data warehouses. Source systems can push events to data warehouse tables via replication software, message queues, change data capture utilities or a combination of the three. These systems can dump data into a real-time ETL adapter, an in-memory cache or some other service that inserts data into the appropriate data warehousing tables.

For example, 1-800 Contacts, mentioned earlier, is currently converting its mini-batch processes to an event-driven system to keep up with growing data volumes and provide greater data integrity, according to Jim Hill, data warehousing manager at the company. Although most business requirements can be satisfied with 15-minute updates, the company plans to deliver an event-driven dashboard in which the data is updated instantaneously as events happen.

OUTSIDE THE DATA WAREHOUSE

- **Operational data store.** Before data warehouses existed, companies tracked data in near real time by querying operational systems directly. This is an inexpensive approach, but it only delivers data from a single system and, if query volumes become too heavy, risks bogging down performance of operational systems. The logical extension is to build an operational data store (ODS) that collects current data from multiple systems and makes it available for real-time queries and operational reporting. But building and maintain an ODS is expensive and requires replicating and moving data. That adds latency to the system, which may not be acceptable for users looking for real-time data.

- **Data virtualization.** A potentially less expensive alternative to an ODS is to federate data using special tools that query multiple systems and join the results on the fly. These data virtualization tools, as they are called now, do a great job of abstracting diverse data sources behind a single "data service" interface, but they are subject to the vicissitudes of source-system performance, network bottlenecks and dirty data. Finally, many BI tools maintain a local cache or an in-memory database to optimize query performance. Often, these caches can be updated incrementally at scheduled intervals, which can be measured in minutes, to ensure users have the most current data possible.

- **Complex event processing.** Another promising approach is complex event processing (CEP). These systems capture, filter, and correlate events emanating from one or more operational systems. These rules-driven systems are like intelligent sensors that organizations can attach to their transaction data to watch for meaningful combinations of events or trends. In essence, CEP systems are sophisticated notification systems designed to monitor real-time events.

Complex event processing systems are sophisticated notification systems designed to monitor real-time events.

Business analysts design CEP rules that correlate data, update dashboards and trigger alerts and actions. Operational workers use CEP dashboards to monitor and manage continuous operations, such as supply chains, transportation operations, factory floors, casinos, hospital emergency rooms, Web-based gaming systems and customer contact centers, among other things, according to Roy Schulte, distinguished analyst at Gartner Inc.

According to Schulte, rules can be designed to 1) capture relevant events from the network; 2) calculate totals, averages, and other statistics; 3) identify

[5]See the book *Event Processing: Designing IT Systems for Agile Companies* by K. Mani Chandy and W. Roy Schulte, McGraw Hill, 2010.

patterns in base events, including trends, particular sequences of events, and causal relationships among events; and 4) enrich the data through comparisons with historical data or augmentation.[5] He believes CEP is in its early adopter phase and will become more commonplace in the next five to 10 years. He also believes many more systems, both transactional and analytical, will be designed as event-driven systems in the near future.

- **Streaming data.** Closely related to CEP systems are stream-based systems that process extremely large volumes of homogeneous event data. They are ideal for handling machine-generated events coming from a single type of device or system, such as discrete events emitted by sensors, although some are designed to process continuous data streaming from audio and video systems. Unlike CEP systems, most streaming systems don't embed sophisticated rules to do correlations and pattern matching, although this is changing as several streaming vendors have acquired CEP vendors to blend the best of both worlds.

- **Early adopter phase.** Schulte believes that traditional BI professionals will be critical in architecting event-driven analytical systems because they have gained fluency in gathering detailed requirements talking with businesspeople, and helping businesspeople weigh the costs and benefits of adopting data-driven systems. Most BI leaders have already traveled partway down the path to continuous intelligence and gained valuable experience. However, the most astute BI leaders won't wait for the business to ask for near-real-time information; they will architect it into their systems so they are ready when the business asks. ■

Content Intelligence

THE LAST DOMAIN of intelligence is perhaps the most important. For two decades business intelligence has only focused on numeric data housed in relational and other database management systems. It has largely ignored the vast majority of data contained in content management systems, documents, email messages, Web pages, social networking sites and electronic data interchange, among others. At best, it has represented unstructured data as binary large objects, known also as blobs, in tables that point to files outside the data warehouse.

Unstructured data often sheds valuable insight about the context of the activity that is represented numerically in relational databases. For example, a content intelligence system might enable a company to correlate sales with customer sentiment about its products. It can help companies better monitor the problems that customers experience with their products and services and even detect fraudulent activity.

Today, there are three approaches to integrate unstructured and structured data:

1. Search with BI. In this approach, business users use BI tools and enterprise search tools independently. They use BI tools to explore numeric data in relational databases and enterprise search tools to explore unstructured data in documents and Web pages and manually bring the result sets together in a spreadsheet or as tacit knowledge. For example, the BI director at a major a movie rental company said that business analysts who notice a dip in sales for a particular movie title in their BI systems will then turn to Google to find out if there have been any recent events that might explain the decline. This approach is widespread today, but it's not always efficient; business users may miss key correlations in the data depending on the search strings they use.

2. Search on BI. BI vendors use this approach to make it easier for business users to find report files or create ad hoc queries. BI vendors embed a search engine that indexes the vendor's report files and metadata. By typing key-

SUMMARY

BI DELIVERY FRAMEWORK 2020

BUSINESS INTELLIGENCE AND ANALYTICS INTELLIGENCE

TOP-DOWN VERSUS BOTTOM-UP

CONTINUOUS INTELLIGENCE

CONTENT INTELLIGENCE

CONCLUSION

words into the search box, users can quickly locate relevant reports and other documents in the targeted file system. Some BI tools using linguistic capabilities (i.e., semantic processing) to map search terms to metadata in the target database (i.e., metrics, dimensions and attributes) to generate ad hoc queries. A related approach, called text mining, uses semantic processing to extract entities and relevant phrases from text documents and embed them in a data warehouse schema so users of BI tools can query both structured and unstructured content using SQL.

3. BI on search. Here, search vendors index both structured and unstructured data, enabling users to explore 360-degree relationships in the data using keyword search and faceted navigation. The tools typically parse entities and concepts from documents using semantic technology and insert them into tables, where they can be associated with related terms and interrogated by users using various techniques. BI-on-search vendors use many different approaches to unify structured and unstructured data. Some, such as Splunk Inc., flatten (i.e., denormalize) relational data into wide tables used in a traditional search-based inverted index. Others, such as Attivio Inc., create hybrid indexes that support both search and SQL queries. Others, such as Endeca Technologies Inc., apply multiple types of indexes to a columnar database that houses both structured and unstructured data.

The third option above is commonly referred to as unified information access (UIA). Its intuitive interface and flexible architecture make it suitable as a complement or replacement for existing BI applications, including data exploration, dashboards and reporting.

BENEFITS OF UIA

- **User interface.** UIA technology offers many benefits to traditional BI users. First and foremost, it has a super easy interface for querying information. It uses a keyword search interface popularized by Google to generate ad hoc queries against structured and unstructured data. To increase the accuracy of search-based queries, these tools use fuzzy matching to adjust for spelling errors and semantic technology to infer the meaning of words that users type into a keyword search box. The search interface is so simple and familiar to so many users today that it will likely become the preferred method of both casual and power users for submitting ad hoc queries. Some UIA platforms also support SQL for more precise and complex queries.

- **Faceted search.** In addition, UIA tools dynamically generate topical categories (i.e., taxonomies or facets) related to the search terms that enable users to refine their queries or explore associations in the data they may not have known existed. This faceted navigation is far more flexible than traditional BI navigation, which is largely limited to predefined drill paths or semantic layers in top-down BI or the skill of the analyst to navigate database schema in bottom-up BI.

- **Flexible schema design.** Perhaps the biggest advantage of UIA tools is that they don't require a rigid schema or fixed field types. Fields can be any length and contain any type of data. One table can contain relational data and the next PowerPoint files. Consequently, administrators don't have to create a schema up-front before they load data into the index. This design provides great flexibility and agility, making it quick and easy to build and modify search-based applications. In many ways, BI on search is a kindred spirit to in-memory visual analysis tools and Hadoop, discussed earlier, which also advocate a schema-less or schema-light design.

Many search engines create an index of some sort—often similar to key-value pair tables—which associates entities (e.g., words or values) with containers (e.g., documents or database tables.) For example, when a user types a word into a keyword search box, the search engines returns links to all documents that contain the word, ranked by relevance or some other criteria based on attributes captured by the engine. Search engines can index any data using this approach, including relational databases. So, administrators can add a new source without having to remodel the structure of the index. The engine simply associates new entities with new or existing containers in an agnostic way. Many UIA tools have a proprietary query language that can filter, join and calculate data in unique ways, often to support the creation of an application, dashboard or report as well as one-off queries.

- **Near-real-time updates.** Search engines are capable of indexing large volumes of data very quickly, depending on the amount of pre-processing that is done, and thus can be used to support continuous intelligence applications. UIA tools can ingest a data feed from a source system and index the data in real time. Many search tools incrementally update the index rather than refresh it from scratch every time there is a change, accelerating data access.

One U.S.-based public media company uses UIA technology to monitor usage of its content management system (CMS). "We've had some amusing situations where a developer sees a user struggling with the CMS and walks over to the person's desk and asks whether they can help, and the person freaks out!" says a systems analyst.

Data integration. In addition, UIA tools can be used as virtual data integration tools. Rather than consolidate data from multiple systems into a data warehouse at great expense, UIA tools leave the data in place and logically integrate the data via metadata (i.e., the search index.) For example, the Canadian Department of National Defence used a BI-on-search tool to analyze cost data from three disparate payroll systems that it didn't have the time or money to physically integrate. Most search queries can be answered by the metadata alone (e.g., a list of the most relevant documents that contain a word or the sum of sales in an order_amount column), but will also contains URL links to the original record (i.e., the document or file) if needed.

SWEET SPOTS AND CHALLENGES

Because of their intuitive interface and flexible architecture, UIA tools make good exploratory analysis tools. They are also well-suited to collecting and displaying related items across structured and unstructured data sources, and performing straightforward calculations on those items. As such, they are suitable for supporting dashboards that track historical or near-real-time data.

However, many UIA tools struggle to model complex dimensions and hierarchies and perform complex calculations against them. Most don't support SQL and so have a difficult time complementing organization's existing BI tools and applications. Security is also challenging to build into indexes without slowing down performance or creating clumsy workarounds. However, some UIA tools have already addressed some or all of these issues and are proving themselves as valuable components within a BI stack. ■

Conclusion

- **Four intelligences.** It is not easy to turn data into information and information into insights and action. For too long, companies have tried to shoehorn all reporting and analysis tasks into a single architecture. This report argues that in the coming decade BI teams will need to embrace four intelligences to deliver insights and action and fulfill the promise of business intelligence.

- **Balance top-down and bottom-up.** First, companies will need to balance top-down and bottom-up approaches to business intelligence. Each has their place. Top-down BI delivers reports and dashboards to casual users who need to monitor performance of key business processes. Bottom-up analytics uses ad hoc tools to answer unanticipated questions and evaluate new proposals and projects. Each approach requires a different architecture: Top-down uses a classic data warehousing architecture, while bottom-up uses analytic sandboxes to ensure ad hoc queries don't proliferate into spreadmarts that undermine information consistency.

- **Super-user network.** While each approach has its place, it's also important to bridge top-down and bottom-up BI with self-service techniques. First and foremost, organizations must bridge these domains manually by implementing a robust super-user network that serves three fundamental purposes: 1) addresses ad hoc information needs of casual users by offloading custom report creation from the corporate BI team; 2) extends the corporate BI team with BI-savvy business users who can help create standard departmental reports and dashboards; and 3) serve on a BI report review board to review requests for new standard departmental reports and dashboards.

- **Self-service BI.** It's also important to implement self-service BI tools for both casual and power users that expose functionality on demand. It's also critical to implement both top-down self-service BI tools for power users (semantic layers and mashups) and bottom-up self-service BI tools for casual users (BI search and visual analysis tools.) At the same time, to fully bridge the worlds of top-down and bottom-up BI, organizations will need to accelerate the ability of their data warehouses to adapt to change. Otherwise, bottom-up approaches will gain ascendancy, creating silos of information that do not align.

■ **New intelligences.** Finally, it's important for BI teams to begin exploring ways to implement continuous intelligence and content intelligence technologies. Many BI teams have already accelerated the refresh intervals of their data warehousing updates, but they may want to explore going the next step and implementing event-driven analytic applications that will enable their organizations to compete on velocity. They should also explore new BI-on-search tools that unify access to both structured and unstructured data through a highly intuitive search interface. ■