WHITE PAPER

Laying the Foundation for Successful Business Analytics Solution Deployments: The Critical Role of Hardware Infrastructure

Sponsored by: IBM

Jean S. Bozman Dan Vesset

December 2012

IDC OPINION

There is growing quantifiable evidence that organizations with higher business analytics competency outperform their less analytically oriented peers. IDC research shows that most analytically oriented organizations are 20% more likely to be among the most competitive organizations within their industry.

The convergence of intelligent devices, social networking, pervasive broadband communications, and analytics is ushering in a new economic era that is redefining relationships among producers, distributors, and consumers of goods and services. In this *intelligent economy*, access to information, combined with the ability to analyze and act upon that information, creates competitive advantage in commercial transactions, enables sustainable management of communities, and promotes appropriate distribution of social, healthcare, and educational services.

Manufacturing, healthcare, telecommunications, public sector, and other organizations that have higher competency and pervasiveness of business analytics solutions are defined by their approach to the selection, deployment, ongoing development, and management of new systems. These organizations not only focus on the software functionality for information integration, monitoring, management, analysis, and visualization but also innovate with regard to the hardware infrastructure that enables successful business analytics projects and ongoing programs. They place a premium on optimization and resiliency and on the ability to scale business analytics, server, and storage infrastructure to address expanding volumes of multistructured data, users, and use cases.

These organizations see a compounding of benefits, with initial project successes setting the foundation for ongoing expanded use of analytics throughout their organizations and in collaboration with their external stakeholders.

Organizations should study the "best practices" that are employed by market leaders and seek to emulate them. These examples should inform new business analytics projects, while providing examples and "lessons learned" should help the project leaders for new analytics projects avoid the pitfalls of projects that were not successful.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

SITUATION OVERVIEW

The volumes of information about business analytics technology, solutions, and best practices frequently neglect to highlight the impact of hardware infrastructure on the success of business analytics projects.

The assumption is that business analytics represents a single, homogeneous, enterprisewide requirement. This assumption leads to many market misconceptions that result in suboptimal system performance, rigid architecture, and costly maintenance — in other words, failed projects. The reality is that:

- Business analytics is an umbrella term that federates multiple related workloads, end-user decision support and automation requirements, and hardware and software technologies.
- A combination of scale-out and scale-up server and storage infrastructure may support data collection and analysis. No one approach will address all use cases.
- Enterprise customers must consider how best to support complex workflows with a range of server and storage deployments. For example, the placement of IT resources often influences the latency and performance of the overall end-to-end workload. In addition, the network infrastructure and its capabilities have a strong bearing on the outcome of server-based and storage-based technologies that support analytics workloads.

Designing Business Analytics Infrastructure

Let's dig one layer deeper in this overview of what works and what doesn't work in business analytics.

Figure 1 illustrates IDC's Decision Management Framework, which helps us visualize the range of business analytics workloads. This framework depicts three decision types and four primary dimensions of each decision type, as shown on each of the four axes in the figure.

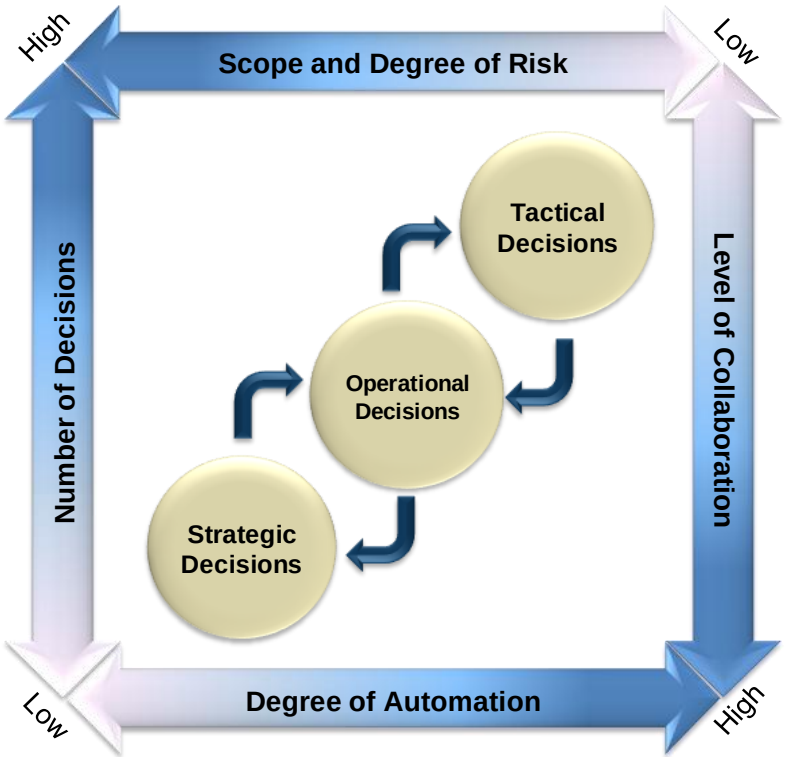
Each of the four primary decision dimensions will determine how an organization will need to address that dimension with hardware and software that make the business analytics solution effective — in terms of functionality and performance as well as initial costs and total costs of ownership.

Business Analytics Requirements

Strategic decisions have the highest scope and risk because of their long-term horizon and the presence of many unknown factors. Although there are relatively few strategic decisions, they require a high level of collaboration among internal and external decision makers, and they allow for relatively little automation.

FIGURE 1

IDC's Decision Management Framework



Source: IDC, 2012

At the other end of the spectrum are tactical decisions made by either front-line employees or by IT systems. There are a lot of these decisions within a given time period, and each decision carries with it relatively little risk and is amenable to automation. Because these decisions are made in the field, in the flow of action, often relatively little collaboration happens during the decision-making process.

Operational decisions lie in between these two endpoints of IDC's Decision Management Framework. Each decision type is associated with different user groups. Operational decisions are made by analysts and managers, strategic decisions are made by executives, and tactical decisions are made by front-line employees or by automated systems, applications, or machines.

Understanding the Decision-Making Process

Understanding your organizational decision-making requirements is an important step toward creating a business analytics strategy for your organization. Importantly, this strategy accounts for all relevant hardware and software considerations, in addition to staffing, financing, and business process considerations. Consider the mapping of typical end-user requirements for each decision type and user group (see Table 1).

TABLE 1

Business Analytics Use Cases and Functional Requirements by Decision Type

Decision Type	Strategic	Operational	Tactical
User group(s)	Executives, managers	Data scientists, quantitative analysts, business analysts	Front-line employees, automated applications
Use cases and functionality requirements	• Risk management • Simulation • KPIs on dashboards • Collaboration • Strategy and performance management • Mobile information access	• Ad hoc analysis • Discovery • Predictive analytics • N-dimensional analysis • Experimentation • Root cause analysis • Collaboration • Management of large amounts of multistructured, multisourced data • Analytic applications focused on planning and discovery	• Structured operational reporting • Application embedded insight • Guided navigation • Rules-based monitoring and execution • Streaming, high-velocity data management • Mobile deployment

Source: IDC, 2012

Leading adopters of successful business analytics solutions have created separate internal groups focused on information management, analytics, infrastructure management, and application development. These organizations have recognized that in addition to specific end-user requirements, common requirements are placed on IT groups, including requirements for rapid deployment of analytic applications, data marts, reports and dashboards, flexible scalability, security administration, and performance optimization that do not represent a drag on business processes.

Importance of Infrastructure Considerations to Successful Business Analytics Deployments

When the "pool" of data was primarily online transaction processing (OLTP), then the generation of 1s and 0s was the main focus of any IT project, and boosting performance was the main driver of new innovation for that project.

Today, no one person or business department can absorb and analyze all of the data that is being generated. In fact, multiple sources of data are giving IT organizations something to think about: engineering data, healthcare data, transportation/logistics data, and — most noticeable — social media data generated by Web sites and mobile phones. So, new approaches must be developed to gather the multistructured data, to store the data, and to analyze the data in a timely way.

Three key infrastructure competencies are found in leading organizations that are able to gain competitive advantage from business analytics: optimization, resiliency, and scalability.

Optimization for Analytics Workloads

- Optimization is the process of improving the performance of a given system by adapting the software for the specific hardware configuration that is deployed. The amount of processing power on the system can be tuned to match a specific workload — such as determining the ideal number of cores and sockets to be housed on the system. Operating systems can be adjusted to run faster on server hardware, and those adjustments can be tested, most often by vendors or IT specialists, to ensure faster processing. Storage and memory can be matched to the specific hardware configuration to ensure that the system is balanced — meaning that there is enough memory and storage to match the processing power aboard the server.

  Compute power is important, but it is not the only consideration for performance. Server I/O, network capacity, and storage characteristics all play important roles in supporting overall processing capacity for analytics workloads. Wherever large amounts of data must be moved from other storage resources, the amount of built-in server I/O capacity to move data can become a limiting factor. So, the number of I/O ports (e.g., PCIe interconnects) and the types of server I/O connections selected are key to this ability to expand the server capacity on the datacenter rack.

  The ability to optimize the system for a particular workload enables a technology vendor to develop and market products that are faster to deploy and easier to maintain. For example, a preconfigured, tested, and certified system decreases the need to install software on the given hardware platform.

- The business analytics technology evaluation and purchasing process depends on several variables. The following are some of the many questions that organizations need to answer:

  - What is the long-term strategic vision for the organization's business analytics platform? Should it include a single, centralized enterprise data warehouse or distributed data marts, or a combination of both?
  - What is the state of the organization's current business analytics architecture and technology? How is it performing? Does it fulfill the needs of all relevant decision-maker user groups?
  - What are the current and expected data volumes and number of users with direct access to the business analytics solution?
  - How frequently does the data need to be captured and loaded into a data warehouse?
  - What is the range of query workloads to be supported by the data warehouse?
  - What types of internal IT and information management skills exist? What skills can a systems integrator provide? What is the right balance between internal and external IT staff?

Depending on the responses to these and other related questions, organizations will have several options for business analytics technology and, specifically, for data warehouses or analytic data stores. Options include the following:

- General-purpose relational database management systems (RDBMSs) that are deployed on a separately purchased hardware platform (including both server and storage technologies). These RDBMSs were developed primarily for transaction processing, but they are also deployed for business analytics.
- General-purpose RDBMSs that are sold without hardware but with specific reference architecture guides for clients that optimize the configuration of the software with various hardware options. These reference configurations are based on tested, certified technology integration between software and hardware but do not include formal product bundling.
- Purpose-built data warehousing or analytics software that is deployed on separately purchased hardware platforms. This is an example of a software appliance.
- Purpose-built data warehousing or analytics software that is preconfigured, optimized, and preinstalled on general-purpose hardware and sold as a single unit by a single vendor. This is an example of a physical business analytics appliance.
- Purpose-built data warehousing or analytics software that is preconfigured, optimized, and preinstalled with purpose-built hardware and sold as a single unit by a single vendor. This is another example of a physical business analytics appliance that represents the highest level of workload optimization.

These five business analytics solution bundles fit into a continuum of functionality in which the first two bundles provide the most flexibility in supporting all of the potential decision support needs, as well as transaction processing needs (regardless of performance). The last three bundles are purpose built for high-performance analytics workloads, but they don't have the flexibility to support certain operational and read/write-intensive business analytics workloads on the same platform.

Resiliency

- **Availability.** For production systems, enterprise workloads must be highly available because they support hundreds, or even thousands, of end users accessing key systems. Although some analytic applications may be used by small groups of data scientists and business unit analysts, availability is still important to overall business success as the role of these analysts has become increasingly critical in delivering insight to other decision makers throughout the organization and to external stakeholders. In a recent IDC survey of over 2,500 organizations, 22% of respondents indicated that if their business analytics solution were out of service for up to one hour, it would have a material negative impact on business operations. That is why IT groups are increasingly paying attention to the resiliency of the systems themselves and to ensuring that these systems are more expandable than expendable — expandable in the sense of addressing more storage, supporting faster I/O interconnects, and being able to analyze ever larger data resources.

- **Flexibility.** Delivered through a more agile infrastructure, flexibility allows IT organizations to better support analytics workloads. Having IT flexibility means that organizations don't need to rip and replace infrastructure. Instead, they are more likely to add net-new servers to existing infrastructure or to reorganize servers into clusters, grids, or arrays that run analytics software. Net-new systems are for new deployments, such as Big Data analytics running the Hadoop algorithms. But, clearly, there is no shortage of either scale-out or scale-up servers that can run analytics software, acting on existing data structures or adding them to available computing resources. One example is deployment of multiple two-socket servers, running analytics software or adding analytics software to scalable servers in the datacenter.

- **Security.** Security is a "must have" feature for databases and analytic applications accessing those databases. Clearly, the data itself must be secure and protected against intrusions from hackers or attackers. However, for analytics to work, applications must reach into secure data repositories and manipulate the data residing in the data stores. Approaches to improve security include the following: isolation of workloads; virtualization of server hardware, where appropriate, to allow workloads to co-reside within a server system; and the use of encryption devices (hardware and firmware) and encryption software.

Scalability

Business analytics tasks can vary in their need for capacity and server performance. For scale-out deployments, each individual server is relatively small and usually an x86 server design. Although each server needs the correct specifications for processors, core count, socket count, memory, and server I/O, clustering software often moves workloads to the server resources they require.

In other cases, data must be imported to a single server for analysis by a business analyst. In those cases, the specifications and requirements are different: The server must be able to import data quickly and efficiently for manipulation and analysis by the analyst. This is also true for the server-based business analytics appliances that are entering the market. Without fast and efficient server I/O, the process of import/export would become cumbersome and impede the workflow to the analyst.

There is a continuing need for scalable servers, including Unix servers and scalable mainframes, that are capable of hosting and processing large amounts of data generated by high-speed transactions that are also running on the same servers. These systems tend to be in the midrange to high-end server classes, and they often have advanced virtualization, including the ability to isolate workloads and to virtualize data in server-controlled storage, and highly granular management capabilities.

Scalability is an attribute that varies from workload to workload and from site to site. There are multiple options to achieve more scalability when it is needed. However, each option depends on a set of hardware and software specifications that allow the computing resources to scale up, as needed, on scale-out or scale-up server deployments.

The evaluation of trade-offs between the various options will depend on the specific business analytics needs of any given organization. IBM provides business analytics technology options that span this full spectrum.

IBM in the Business Analytics Market

IBM provides a broad range of software, hardware, and professional services to address the requirements of the business analytics market. The company's business analytics software portfolio includes tools for data warehousing, data integration, query and reporting, data visualization, predictive analytics, content analytics, spatial information management, and complex event processing.

In recent years, IBM has also bolstered its share of the analytic applications market through acquisitions, including Coremetrics, Algorithmics, Clarity, i2, OpenPages, and others. A sample of the software product brands IBM has assembled under its business analytics umbrella includes InfoSphere, DB2, Cognos, and SPSS. These and related products can be deployed on multiple IBM hardware infrastructure platforms. They are marketed by IBM as business analytics products that enable the Smarter Planet vision laid out by the company.

One of the company's workload-optimized systems is IBM Netezza — an integrated hardware and software appliance purpose built for data warehousing. IBM offers multiple product configurations under the Netezza brand that not only provide the core data management functionality but also support predictive analytics and spatial information management. Netezza is an example of one of the most successful business analytics appliances — both before and after its acquisition by IBM.

IBM's Infrastructure Portfolio for Business Analytics

IBM System x

IBM System x is IBM's x86 server product line that supports all x86 server–based software for virtualization, Big Data, and business analytics. These systems are shipped in a variety of form factors, including rack based and blades, and are the building blocks for x86 server–based infrastructure supporting end-to-end business analytics workloads. They run Linux distributions (Red Hat RHEL and SUSE SLES distributions) and Microsoft Windows server software, along with associated hypervisors (e.g., VMware vSphere, Microsoft Hyper-V, Xen, and KVM).

IBM Power Systems

IBM Power Systems are highly virtualized IBM-designed and IBM-built systems based on IBM POWER multicore processors. They run the PowerVM-branded hypervisor, which offers security, efficiency, and density for Linux (Red Hat RHEL and SUSE SLES distributions), IBM AIX Unix, and IBM i operating environments.

Power Systems deliver scalability and performance for different business analytics and Big Data workloads. Multiple deployment styles are supported for business analytics, including server nodes for business analytics grids and scalable server nodes to analyze data stored in data warehouses or data marts. The proximity of data stored near — or on — these systems enhances overall throughput — as does a portfolio of fast server I/O links to connect to external storage and networking switches.

IBM System z

IBM System z (including the IBM zEnterprise EC12 model introduced in 2012) is a highly virtualized server, via logical partitioning (LPARs), and supports a range of operating systems, including System z's z/OS environment and Linux distributions running on System z Integrated Facility for Linux (IFL) specialty processors. This built-in support for Linux opens the System z to a wide variety of business analytics use cases, supporting workload consolidation onto System z and also linking to other servers (including IBM System x blades or Power Systems blades). System z also has one of the highest security levels available in a commercial server, with the ability to encrypt up to 19,000 data transactions per second. System z is shipping in the flagship zEnterprise 196 models and also in z114 systems, which are priced less than $100,000.

IBM System z with ZBX Blade Chassis

Importantly for business analytics, IBM introduced the ZBX blade chassis as a co-installed solution that supports end-to-end workloads, where the application modules run on the blades and the scalable database resource runs on System z. This solution has a built-in, optimized server I/O that is supported by firmware, which speeds overall throughput from the System z and ZBX combined resources.

IBM DB2 Analytics Accelerator on System z

IBM DB2 Analytics Accelerator is a high-performance software offering that is part of an integrated hardware/software solution designed to work with IBM System z to deliver dramatically faster analytic query responses transparently to users. It integrates into DB2 for z/OS data warehouse environments, forming a high-performance analytic query appliance powered by Netezza technology. The total solution is an integration of IBM hardware, software, storage, and advanced technologies focused on business analytics. DB2 Analytics Accelerator extends System z qualities of service of manageability, security, and availability to analytic applications, working seamlessly with the prerequisite IBM hardware.

IBM PureSystems

In 1H12, IBM took a clean-sheet approach with IBM PureSystems, which is a series of integrated systems supporting choice and flexibility in server deployments, including both x86 and non-x86 server nodes. The nodes are connected via a high-speed fabric and built-in switches from Blade Networks, a company that IBM acquired in 2010. These integrated systems include built-in security features and support flexible deployments, in which nodes can be added later on, as demand for computing resources grows over time.

Shared system resources are managed in a unified way, from a single console, and deployments are based on reference architectures or templates.

Three major areas of focus in the IBM PureScale design are as follows:

- **Scale-in systems design.** The workload becomes the virtual machine or bare metal operating system image, and the system's overall design automates integrated server, storage, and networking resources for both physical and virtual environments.

- **Patterns of expertise.** The systems embed software tools that automate configuration, upgrades, and application requirements.

- **Cloud-ready integration.** IBM PureSystems are designed to be deployed in support of private self-service cloud infrastructure and software to support flexible scale-up and provisioning of workloads.

Importantly, IBM introduced the IBM PureData Systems to the IBM PureSystems family as a way to meet the challenges of growing data volume, variety, and velocity. In the past, IT organizations often purchased general-purpose systems that were not optimized for particular analytics workloads. As a result, they spent valuable time and resources tuning and maintaining these general-purpose systems for the varying requirements associated with transaction and analytics workloads.

To directly address this lack of efficiency, IBM designed and built PureData Systems optimized specifically for the unique demands of transaction and analytics workloads and introduced three PureData offerings in October 2012: PureData Systems for Transactions, PureData Systems for Operational Analytics, and PureData Systems for Analytics. These PureData offerings incorporate the design principles of built-in expertise, simplified and fast deployment, and streamlined management.

IBM Storage Products

The spectrum, or continuum, approach is being applied to analytics because of the variation in processing styles. For this reason, IBM has a portfolio of storage options that can be used for analytics deployments. The Smarter Storage approach, which IBM unveiled earlier this year, offers a way to consider IBM's storage capabilities in the context of analytics deployments. Smarter Storage, IBM's approach to the design and deployment of storage systems, is described by the company as follows:

- **Efficient by design.** Storage systems have built-in capabilities such as IBM Real-time Compression to help organizations manage costs and capacity.

- **Self-optimizing.** Storage systems work automatically and without administrator involvement to improve storage performance and productivity with technologies such as automated tiering (IBM's version is called Easy Tier) and flash-optimized storage.

- **Cloud agile.** Storage systems are deployed using cloud technologies to help increase information access with IBM capabilities such as SmartCloud Virtual Storage Center and Active Cloud Engine for its unified and file products, Storwize V7000 Unified and SONAS.

Customer Case Studies

IDC interviewed four IBM customers that have deployed business analytics solutions based on IBM server and storage products. The sites had different deployment styles and implemented their solutions by taking different approaches to building out the infrastructure that was supporting the business analytics workloads. Some of these clients also used IBM business analytics software, while others relied on third-party software on the IBM hardware infrastructure.

Vestas Wind Systems

Changing a strategic business process is never easy, yet doing so can lead to significant quantifiable profitability benefits. Vestas, a Danish wind turbine manufacturer, has been able to do just that with a new process for selecting sites for its turbines. With customers in more than 70 countries, Vestas manufactures, installs, and operates wind turbines that produce electricity for commercial and consumer uses. Placement of the wind turbines is critical to successful deployments.

Selecting the optimal site for a wind turbine is a high-risk, high-reward decision. In the past, it required Vestas employees to go to a proposed location and to spend up to 15 months measuring the local climate conditions — a process that could cost upwards of $300,000. These direct costs pale in comparison to costs associated with lost opportunity to maximize revenue from a wind turbine that is built in a suboptimal location or from higher ongoing servicing costs for such turbines (e.g., due to excessive turbulence).

Placement of a wind turbine is a complex, multivariable optimization decision that includes considerations about wind turbulence, ecology, and aesthetics — and Vestas was keen to improve the process. To do so, the company deployed IBM InfoSphere BigInsights at its Danish datacenter. In mid-2012, Vestas was running BigInsights software on over 1,300 connected, workload-optimized IBM System x iDataPlex bladed servers that make up the IBM Firestorm supercomputer. Using InfiniBand interconnects, the system has over 2,600 spinning disks and is managed using the Extreme Cloud Administration Toolkit (xCAT), which is an open source distributed computing management software developed by IBM for the deployment and administration of Linux clusters or IBM AIX-based clusters.

Vestas IT executives are seeing the benefit of this approach to analytics-based deployment decisions. Anders Rhod Gregersen, senior specialist, Wind and Site Competence Centre at Vestas, said, "With this new solution from IBM, we are effectively running Hadoop on a Ferrari rather than a tractor."

With the IBM solutions, instead of placing sensors in the field and collecting data for months, Vestas analysts can place a number of virtual wind turbines in specific locations and are able to evaluate the wind trends at those locations for the past 12 years. This system allows analysts to simulate the atmosphere across the entire planet and to draw out over 35,000 measurement points. The data includes multiple petabytes (PB) of structured and unstructured data, such as weather reports, tidal phases, sensor data, satellite images, and deforestation maps.

The results of data modeling provide critical information to Vestas analysts and management within minutes rather than months, as was the case when gathering relevant data before the new system was deployed. While the accuracy of the new data modeling process is similar to the "in the field" data monitoring used in the past, the speed of data processing enables Vestas to do new types of forecasting and deeper analysis. This has led to significant decreases in decision cycles and to increases in the certainty of decisions based on that data. In addition to improving the initial site selection process, the new solution is also being used for predictive maintenance of wind turbines.

WellPoint Inc.

Better healthcare providers, better healthcare equipment, and better access to healthcare are typically listed as the ingredients of improving healthcare for all. What is not always mentioned on this list is better analytics. Yet, leading healthcare organizations are increasingly turning to Big Data and analytics to gain insights that drive decisions to improve all aspects of providing healthcare services.

One of these leading organizations is WellPoint, one of the largest health benefits company in the United States in terms of medical membership, with 34 million members in its affiliated health plans and approximately 65 million individuals served through its subsidiaries. The company is an independent licensee of the Blue Cross and Blue Shield Association. WellPoint was looking to improve its medical review process — the policy-based monitoring, evaluation, and approval process that affects all patient visits and procedures. Previously, the process was manual — requiring days for approvals and reviews — and, importantly, it had been asynchronous with the ongoing work of healthcare providers.

In September 2011, WellPoint announced a partnership with IBM for the application of IBM's Watson technology, initially, for the medical review process. Deployed in December 2011, the technology has accelerated the review of medical procedures and has created a new dialogue between physicians and Watson-supported procedure reviewers. The process now involves physicians submitting patient information and the rationale for requesting specific procedures via an application that transmits this information to Watson. (Readers may recall that Watson technology was showcased on the television show *Jeopardy!* where the technology's analytics and fast processing capabilities allowed the Watson computer to compete with human players and to win games.)

In this application for health services, IBM Watson decomposes the information, compares it against medical guidelines (currently about 200 guidelines and policies) based upon previous medical reviews, and responds to physicians. (Today, the response is sent to the WellPoint nursing staff, which interacts with providers. However, work is under way now to create a direct interface from Watson to physicians; the interface is scheduled to go online in 2013.) Watson's response not only involves the approval or the denial of a given request but also includes the medical evidence that was referenced, and used, to arrive at the decision.

WellPoint is also working with IBM and several large healthcare providers to develop Watson into an oncology diagnosis and treatment solution for cancer cases. In the current pilot project, IBM Watson is ingesting oncology medical evidence, and it is "learning" this new information. Once Watson is trained to review a particular patient's case, the goal is to enable physicians to present particular, new cases to Watson and to receive treatment recommendations that would allow for more personalized, customized treatments.

Other projects utilizing IBM technology include analysis of data to identify patients who might benefit from care management solutions and the creation of a "single source of truth" about provider data. The latter project is based on IBM Initiate master data management (MDM) software, with the goal of rationalizing the highly fragmented provider information. This cleansed data will be hosted and made

available more broadly in the industry, where an estimated 5–10% of claims require manual review due to unclear provider information.

This project shines a spotlight on the ways in which unstructured data and structured data can both coexist and be tapped for analytics solutions. IBM Watson ingests both structured data and unstructured text. The latter includes results of medical image interpretations (noting that ingestion of actual images is also in the plans).

Analytics for healthcare brings with it special requirements for data privacy and security. Protection of patient data is accomplished through various data "anonymization" techniques before the data is consumed by business analysts. In all cases, IBM Watson was described by Rickey Tang, vice president and chief technology officer (CTO), as a stateless system that does not retain any information about the patient because patient data is de-identified as it is presented to Watson.

WellPoint initially hosted the IBM Watson system in its datacenter, but it is moving to a cloud-based solution. A cloud deployment will eliminate the need to keep multiple Watson systems synchronized. Importantly, this approach will enable easier sharing of information with providers and other stakeholders in the healthcare industry.

WellPoint has a large IT organization and utilizes technology from several IT vendors, including multiple products from IBM, such as Watson, DB2, UDB, IMS, IBM WebSphere, Rational, Tivoli, and Netezza, which is used for cost-of-care analysis. As part of the partnership, which also includes a community of physicians providing expertise, WellPoint and IBM developed a lot of enabling technology around the Watson analytic engine. While some of this technology will remain proprietary for use by WellPoint, part of the output of the collaboration will be made available to the rest of the healthcare industry — in a new and emerging business model of leveraging cloud technology investments to recoup development costs through provision of the cloud-enabled analytics services to third-party customers.

IBM Watson has enabled WellPoint to optimize existing processes and to do things that were not feasible in the past, such as dynamic medical reviews and oncology reviews. According to Andrew J. Lang, senior vice president and CIO, the new solution has met with a positive response from doctors, who see it as a decision support tool. The volume of new annual medical information is overwhelming for any human being to absorb, and today, the selection of suboptimal treatment protocols can range as high as 50%. IBM Watson provides targeted medical evidence and, potentially, a range of suggestions while leaving the decision in the hands of the provider. In turn, this increases the chances that the best medical procedures are considered.

AXTEL

The competitive pressures and growth opportunities facing today's telecommunications companies increasingly require adoption of robust business analytics solutions to enable faster and more insightful decision making. Among the companies that have recognized this need for analytics and the benefits it can provide is AXTEL, one of Mexico's largest telecommunications providers. Based in Monterrey, Mexico, the company had more than $775 million in annual revenue in 2011.

To support the information access and reporting requirements of commercial clients, AXTEL used to ship CDs with usage data to each client. Due to its complexity and cost, this unsustainable practice became a drag on customer service and a risk factor in supporting ongoing growth of the company.

To address the issue, AXTEL launched a new business intelligence project focused on proving business customers with faster access to more detailed data about their communications services usage. As a major supplier of services to commercial and government entities, AXTEL provides these stakeholders with direct access to a data warehouse of over 10 terabytes (TB) of usage data. The data warehouse, which is now deployed on servers powered by IBM POWER7 processors and an IBM XIV Storage system, is a critical system that supports operational reporting requirements of a large concurrent number of users.

AXTEL provides clients with full, raw data sets and is in the process of creating value-added, preconfigured reports of the most commonly requested views and metrics. The data is typically kept in the data warehouse up to six months, and then it is archived using IBM's SAN-based storage and archiving technology.

In addition to supporting the external usage reporting function and internal performance management needs, the IBM infrastructure supports a range of operational and analytic applications, including SAP, Oracle Siebel, MicroStrategy, and Informatica.

Gaspar Rivera Del Valle, AXTEL's IT evolution director, said that the latest move to IBM POWER7 infrastructure has resulted in 30–40% data processing performance improvements, with data loads occurring every five minutes. This is much faster than the performance he was seeing with the previous system. The new system has improved customer service, as measured by the reduction in customer complaints. AXTEL's large customers, including customers who are operating call centers, no longer need to load large data sets from CDs in order to access their usage data.

Internally, AXTEL has eliminated the time and expense of creating individual CDs, and management and analysts now have access to vital information in real time (as soon as the information is loaded into the data warehouse). Internal users of the data include AXTEL employees in the finance, marketing, sales, call center, and customer service groups.

Miami-Dade County

Miami-Dade County is a large metropolitan area that has gathered large data repositories, based on transactional data that can be mined for actionable information and made available to business analysts. At the same time, some of the county's data, such as public safety and municipal services data, can also be made available, via Web portals, to citizens.

IBM Cognos business intelligence software enables the public and business analysts to view data extracts that would otherwise have been difficult or time consuming to access on paper records. "Cognos on the mainframe supports transparency in government," said Anita Nolan, senior operating systems specialist at Miami-Dade. "For example, every check that the county has written over the last seven years is on [the IBM] System z." The System z is a high-end, mainframe-class server system.

Indeed, transactions — and protecting the transactional data — is a core value at the Miami Dade site. "We have a secure platform that is available on a 24 x 7 basis," said Jose Eskert, senior operating systems specialist at Miami-Dade. "These systems run for months, and we schedule a planned one-hour or three-hour outage, depending on the maintenance we want to perform, and the systems stay up and running forever." In addition, extensive backup processes run every day to ensure that data is copied in multiple sites, in case it is needed for recovery.

The county decided to leverage its existing base of IBM Cognos software, which was already resident on two IBM System z mainframe systems. The IBM Cognos 8 installation was sizable, with 1,500 named users and 1,600 dashboards. Miami-Dade is planning to double that capacity over a period of 18 months and to grow capacity by adding more Linux "engines" to its existing IBM z10 Business Class systems.

At Miami-Dade, data is housed on two IBM System z10 Business Class servers, and the analytics go against data extracts from the live, production databases. Specifically, Miami-Dade runs IBM Cognos 8 on Linux, which is running on the System z Integrated Facility for Linux (IFL) specialty engines (specialized processors that run Linux). Earlier this year, the county added three IFLs, bringing the total to five IFLs on each System z server. These processors host native Linux software environments, which in turn support a range of Linux-enabled workloads, including business analytics.

Scalability is a consideration for this deployment — as is IT flexibility. Scalability was supported by the addition of more IFLs, but it was also supported by the ability to add more logical partitions (LPARs) to either System z server, when needed. This is a very real example of IBM's virtualization technology, which has been part of the IBM System z product line (and preceding mainframe models) for four decades. New computing resources can be provisioned, as needed, without acquiring new servers or changing datacenter deployments of physical servers.

In addition to the central-site processing on IBM System z, some of the data extracts are hosted on IBM Power Systems Unix servers to bring the data closer to the business analysts and users who are accessing the data. These deployments support important initiatives on the countywide level, as in the following examples. Municipal services data can now be accessed by county residents, via the Web — such as for police, fire, water, and sanitation services. Another application focuses on energy conservation in government buildings and provides feedback to county employees on the amount of energy use occurring in their department offices. Reports on prisons and jails are automatically generated now — a process that used to take a week for the court system to complete. This approach reduces the total number of phone calls to city services personnel — and thus has a direct impact on operational costs.

This Linux on System z technology also has the effect of reducing IT staff costs because the number of support staff who are needed for the mainframe has not increased, even though the workload for IBM Cognos business analytics software has been added to the System z environment, by running in logical partitions (LPARs) on the System z server. Only four IT staffers support this Linux/System z environment, which is being accessed by about 28,000 Miami-Dade County employees.

CHALLENGES/OPPORTUNITIES

Like any rapidly growing market opportunity, the business analytics market presents many challenges to those who would seek to harness the power of analytics for efficiency, innovation, or control. The priorities for all analytic workloads are as follows: Organizational goals need to be understood, user requirements need to be defined; data sources and types need to be identified; the right tools for IT and business unit teams need to be selected; and ongoing programs need to be established to continuously re-evaluate all of the preceding factors.

IBM has a broad portfolio of market-tested products and services to address business analytics requirements. These offerings include infrastructure and software that have been optimized to support analytics workloads. However, IBM will clearly be competing with other large companies that see the same opportunity -- and with a range of smaller companies that can work to disrupt the "status quo" and to up-end the traditional business with new technologies and approaches to analytics. IBM's ultimate success in the business analytics market will result from its understanding of both the opportunities and the issues for customers. IBM's success will also depend on the company's ability to field useful business analytics software products, over time, and a range of appropriate server and storage infrastructure.

IBM's ability to follow up with services, support, and consulting to assist, support, and guide its customers worldwide will also be key. IDC believes that IBM will continue to invest deeply in analytics — and to look to a combination of organic growth and acquisitions to maintain its momentum in the dynamic and rapidly growing business analytics market.

RECOMMENDATIONS

Infrastructure cannot — and should not — be an afterthought. As customers who have adopted business analytics can attest, the most flexible, scalable, and resilient analytics systems have been developed — and put into production — through thoughtful implementations. Organizations across the intelligent economy now being deployed around the world should consider the following best practices:

- Develop a business analytics strategy (or review an existing business analytics strategy). Utilize IDC's Decision Management Framework (refer back to Figure 1) to address strategy components such as decision types, decision makers, metrics and key performance indicators (KPIs) information latency requirements, data sources and data types, and technology and services. However, having a business analytics strategy does not require enterprisewide, large-scale project deployments.
- Recognize that for business analytics, one size (technology) does not fit all (requirements). Different workloads, data types, and user types are best served by technology that is purpose built for a specific use case. Hadoop may be the best choice for large-scale Web log analysis, but it certainly is not the technology of choice to analyze real-time streaming data. An MPP RDBMS is great at structured analysis of large amounts of operational data, while an in-memory analytic engine would be best suited for rapid, *ad hoc* evaluation of alternative scenarios. In each case, there is an opportunity to deploy hardware infrastructure that is optimized for the specific software and the use case.

- Execute incrementally. Most success stories in the business analytics market are defined by well-scoped and well-defined projects that allow the business analytics team to demonstrate rapid ROI. Success leads to additional projects and eventually pervasive use of business analytics capabilities throughout the organization.

- Determine the project's need for hardware infrastructure and software requirements in parallel. Although business users are likely to provide most of their input to the software requirements, IT groups must ensure that hardware infrastructure selection does not become an afterthought. For example, if end users require real-time access to data to perform rapid scenario evaluation or *ad hoc* planning, they will need to be supported by an in-memory system. If they require access to dozens or hundreds of TB, or even PB, of data, storage requirements must be able to accommodate such a volume of data.

- Consider the performance impact of enabling hardware infrastructure. Choices such as whether to use in-memory computing or MPP analytic databases, appliances, or separate components will have a material impact on the success of the customer's business analytics solution. A strategy should be developed that defines how to achieve goals for scalability, resiliency, and availability of all parts of the system.

- Carefully evaluate and repeatedly test the "feeds and speeds" of the analytics infrastructure as the project scales to provide more capacity — in terms of both the size of the system infrastructure and the amount of data to be analyzed or the number of users gaining access to the data. Key considerations include capacity of servers and storage devices, amount and size of data caches for temporary housing of data, and latency associated with passing the data from one server node to another or from a storage device back to the servers for processing. Once the multi-TB range is reached, consideration must be given to the practices of deduplication of redundant data — and to checking the integrity of the data being analyzed. Any corruption in that data will invalidate the results of the analysis.

- Enable the highest value-added work for all employees. In the context of business analytics, this means ensuring an appropriate balance of centralized functions performance by IT and self-service functionality provided to business end users. Understanding which business intelligence (BI) and analytics tasks can be centrally "productized" to support repeatable information access and analysis requirements and which tasks need to support *ad hoc* exploratory data analysis will determine the infrastructure, software, and staffing requirements for the various components of the overall business analytics solution.

- Don't lose sight of the goals set forth in your business analytics strategy. Many IT and information management groups become enamored with the pursuit of technical excellence and lose sight of the business goals. The "if we build it, they will come" approach to business analytics rarely succeeds. At the same time, IT teams need to be ready to respond rapidly to the inherently ever-changing business analytics requirements of business users.

- Consider that there is more to business analytics than technology. The best infrastructure, the best data, and the best analysts and analysis don't guarantee business success. The organization must ensure that the output of any business analytics solution contributes to the decision-making processes of all decision makers.

CONCLUSION

Recognize that the goal of any business analytics project is not the collection of a large amount of data or the employment of a virtual army of data scientists. Business analytics projects must support concrete business needs, and they must provide actionable information to decision makers, whether they are executives, line-of-business (LOB) employees, or automated systems.

Yet business analytics technology matters. Optimization, resiliency, and scalability of business analytics solutions can be the difference between making a timely decision based on latest information or blindly following your gut. In the intelligent economy, which is full of large volumes of high-velocity, multistructured data, the latter approach to decision making is no longer an option.

Flexibility in the design of analytics infrastructure is a key consideration for any new analytics project. IT groups as well as business users will benefit by recognizing that deploying appropriate infrastructure to support a wide range of business analytics workloads will require constant evaluation and a willingness to adjust the infrastructure as needed. Ultimately, the responsiveness of the resulting systems is highly important to the success of analytics projects. The amount of time it takes end users to find their business "answers" is key to a project's success and to business users' perception of the quality of the internal IT group's performance.