Driving Business Optimization with Trusted Information

Leveraging IBM InfoSphere and Cognos Solutions

Learn about the latest innovations from IBM Information on Demand:

- Information Agenda
- InfoSphere Software
- Cognos

Driving Business Optimization with Trusted Information

Leveraging IBM InfoSphere and Cognos Solutions

Bill Wong, Dr. Guenter Sauter, Brian Byrne, Dan Wolfson,
Harriet Fryman, Paulo Pereira, Dr. William O'Connell,
Phil Downey, Rex Wiederanders, Alan Meyer

MC Press Online, LP

Lewisville, TX 75077

Driving Business Optimization with Trusted Information
Leveraging IBM InfoSphere and Cognos Solutions
Bill Wong, Dr. Guenter Sauter, Brian Byrne, Dan Wolfson, Harriet Fryman, Paulo Pereira, Dr. William O'Connell, Phil Downey, Rex Wiederanders, Alan Meyer

First Edition
First Printing—October 2008

For information regarding permissions or special orders, please contact:

MC Press
Corporate Offices
125 N. Woodland Trail
Lewisville, TX 75077 USA

For information regarding sales and/or customer service, please contact:

MC Press
P.O. Box 4300
Big Sandy, TX 75755-4300 USA

ISBN: 978-158347-088-6

Acknowledgements

We would like to thank and acknowledge the following people for their assistance on and guidance of this primer: Joseph Labriola, Gaurav Deshpande, Marcelo Cabane, Susan Visser, Lynn Jonas, Eddie Daghelian, Don Cambpell, Jacqueline Coolidge, Meg Dussault, Tom Inman, Inhi Cho, Greg Lotko, Arvind Krishna, Ambuj Goyal, and Steve Mills .

About the Authors

Bill Wong is the InfoSphere executive responsible for delivering solutions built on IBM InfoSphere Warehouse and InfoSphere Balanced Warehouse. He works extensively with customers, consultants, and analysts to position IBM's current and future offerings. He has supported the IBM data warehouse offerings since their introduction, with roles in development, marketing, and sales. Bill has spent several years assisting customers implement solutions in the financial services, retail, and telecommunications industries, and his previous book dealt with business intelligence and warehouse solutions in the life sciences industry.

Dr. Guenter Sauter is a senior architect at IBM working on trusted information, information integration, master data management, and information as a service. He is a member of IBM's patterns governance board and the technical lead for IBM's SOA Scenario on Information as a Service. Before joining IBM in 2004, Guenter spent 10 years in the automotive industry in various IT architecture and management roles. He holds a Ph.D. in computer science from the Technical University of Kaiserslautern, Germany.

Brian Byrne is a cross-product architect at IBM, addressing IBM Information Server, Cognos, MDM Server, and Industry Models. Brian has extensive experience working with customers to deliver information as a service and SOA-based solutions in financial services and has a 10-year career in the architecture, design, and development of distributed software systems.

Dan Wolfson is an IBM Distinguished Engineer and the chief architect and CTO for the Information Platform and Solutions segment of the IBM Information Management Division of the IBM Software Group. He is responsible for architecture and technical leadership across the rapidly growing areas of information integration, master data management, and industry models. Dan's previous roles include CTO for Business Integration Software and chief architect for Information Integration Solutions.

Harriet Fryman is the product marketing manager for the Cognos business intelligence and performance management platform and currently leads the market team to drive Cognos synergy with the broader IBM portfolio. Harriet has 20 years' experience in BI, data warehousing, OLAP, and data integration, including eight years hands-on with customer implementations. She is a spokesperson and writer on key

BI trends, including how to drive greater user adoption and engage businesses in a two-way dialogue with their information to drive business performance.

Paulo Pereira has 18 years of experience in the field of information management, having spent most of his time working directly with clients and helping drive technology developments that would benefit them. Formerly the chief data warehousing technologist for the Worldwide BI Best Practices team and the chief Information On Demand architect for the IBM Software Group, Paulo collaborates with IBM labs, partners, and services organizations in projects that help define IBM Software Group's strategy around data warehousing and Information On Demand.

Dr. William O'Connell is an IBM Distinguished Engineer and the CTO for the Data Warehousing Solutions segment of the IBM Information Management Division of the IBM Software Group. He is responsible for architecture and technical leadership across the InfoSphere Warehousing product space, including the core database, tooling, and analytics and InfoSphere Balanced Warehouse.

Phil Downey is the technical marketing program manager for IBM InfoSphere Warehousing and is part of a team driving product direction and development of IBM's leading warehousing solutions. Phil has been working in the areas of data warehousing and business intelligence for the past 15 years and has broad experience across a variety of industries and implementations, both as part of an independent Systems Integrator and in later years with IBM.

Rex Wiederanders is manager of the Worldwide Technical Sales Enablement group, of which he has been a member for 11 years. Rex has held various positions in IBM since 1981, including sales, technical sales, development, and above region specialties. Concentrating on software for the past 20 years, Rex has been a frequent speaker at many trade shows and user groups, where his presentation style has been well-received.

Alan Meyer is the marketing manager for System z Data Integration in IBM's InfoSphere organization. He has been involved in the BI community since the mid-1980s and has held both technical and marketing positions in data warehousing and business intelligence. Alan has been a frequent speaker at industry trade shows and has written numerous papers on topics of innovating with information.

Contents

Foreword

Welcome and thank you for taking time to understand the latest evolution from the IBM Information on Demand strategy.

IBM Information on Demand helps our customers unlock the value of information to optimize business performance. In the past, technology investments have focused on business automation driven by an application agenda to drive business value. Companies are now investing in business optimization efforts as well, creating an Information Agenda to transform their information into a trusted strategic asset for sustained competitive advantage.

Business optimization requires information that is accurate, complete, in context, and actionable. Achieving this level of trusted business information requires transforming, reconciling, and maintaining information, and delivering it in real time to the people, processes, and applications that need it. IBM InfoSphere and Cognos lead the industry by offering the breadth of capabilities required for the end to end management and delivery of trusted business information to ensure competitive success.

IBM is committed to your success and we have assembled our best and most sought after subject matter experts to deliver this introductory primer to you. We offer this primer as a guide for your business optimization efforts.

Inho Cho
Vice President
Data Management Marketing

Greg Lotko
Vice President
Warehouse Solutions

1

Information as a Strategic Asset

What will the "Enterprise of the Future" look like? To answer that question, we spoke with more than 1,000 CEOs from around the world as part of IBM's 2008 Global CEO Study. These conversations, together with our statistical and financial analyses, provide a unique perspective on the future of the enterprise.

Coping with New Kinds of Change

Until just recently, market factors — such as customer trends, market shifts, and competitors' actions — dominated the CEO agenda. Other external factors — socio-economic, geopolitical, and environmental issues — were seen as less critical, rarely making it to the CEO's desk. Today, the priority list of the CEO has expanded, resulting in increased risk and uncertainty. People skills are now just as much in focus as market factors, and environmental issues demand twice as much attention as they did in the past. Suddenly, everything is important. And change can come from anywhere.

CEOs are most concerned about the impact of three external forces they anticipate will cause the most change for their organizations over the next three years: market factors, people skills, and technology. Shifts in customer expectations, competitive threats, and industry consolidation continue to weigh on their minds. CEOs are also searching for industry, technical, and particularly management skills to support geographic expansion and to replace aging baby boomers who are exiting the workforce. CEOs in our survey rated insufficient talent as the top barrier to global integration — even higher than regulatory and budgetary hurdles.

CEOs also described how technological advances are reshaping value chains, influencing products and services, and changing the way their companies interact with customers. CEOs told us they're changing their business models because they are finding it increasingly difficult to differentiate based on products and services alone. But they also stressed another reason: they simply have more options now.

Today's CEOs are rapidly positioning their businesses to capture the growth opportunities they see. Our discussions about their plans and challenges revealed several striking findings:

- Organizations are bombarded by change, and many are struggling to keep up. Eight out of 10 CEOs see significant change ahead, yet the gap between expected change and the ability to manage it has almost tripled since our last Global CEO Study in 2006.
- CEOs view more-demanding customers not as a threat, but as an opportunity to differentiate. Businesses are spending more to attract and retain increasingly prosperous, informed, and socially aware customers.
- Nearly all CEOs are adapting their business models; two-thirds are implementing extensive innovations. More than 40 percent are changing their enterprise models to be more collaborative.
- CEOs are moving aggressively toward global business designs, deeply changing their organization's capabilities and partnering more extensively. They have moved beyond the cliché of globalization, and organizations of all sizes are reconfiguring to take advantage of global integration opportunities.
- Financial outperformers are making bolder plays. These companies anticipate more change and manage it better. They are also more global in their business designs, partner more extensively, and choose more disruptive forms of business-model innovation.

Introducing the Enterprise of the Future

These findings — across industries, geographies, and organizations of different sizes — paint a surprisingly similar view of the qualities we believe will be needed for future success. At its core, the Enterprise of the Future possesses the five traits shown in Figure 1.1.

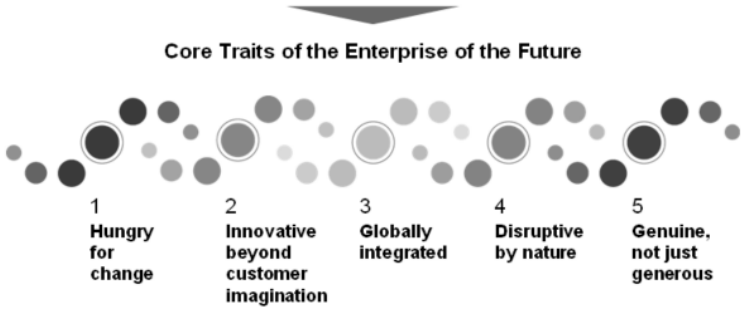

Figure 1.1: Core traits of the successful enterprise of the future

- *Hungry for change* — The Enterprise of the Future is capable of changing quickly and successfully. Instead of merely responding to trends, it shapes

and leads them. Market and industry shifts are viewed as opportunities to move ahead of the competition.

- *Innovative beyond customer imagination* — The Enterprise of the Future surpasses the expectations of increasingly demanding customers. Deep collaborative relationships enable it to surprise customers with innovations that make both its customers and its own business more successful.
- *Globally integrated* — The Enterprise of the Future is integrating to take advantage of today's global economy. Its business is strategically designed to access the best capabilities, knowledge, and assets from wherever they reside in the world and apply them wherever they are required in the world.
- *Disruptive by nature* — The Enterprise of the Future radically challenges its business model, disrupting the basis of competition. It shifts the value proposition and overturns traditional delivery approaches. As soon as opportunities arise, it reinvents itself and its entire industry.
- *Genuine, not just generous* — The Enterprise of the Future goes beyond philanthropy and compliance and reflects genuine concern for society in all its actions and decisions.

Business Strategy Implications

The Enterprise of the Future aims beyond articulated needs and wants, creating first-of-a-kind products, services, and experiences that were never asked for but are precisely what customers desire. The enterprise achieves this type of success in a variety of ways.

The Enterprise of the Future *finds ways to make offerings relevant to new markets and to increasingly prosperous consumers*. Global brands, products, and services deliver economies of scale, yet each market has its own culture, needs, and aspirations. The Enterprise of the Future constantly experiments and learns how to optimize the balance. It analyzes potential markets to find niches, "white space," and complacent competitors where it can capitalize on its core strengths.

The Enterprise of the Future *understands timing and network effects*. There is a fine line between "beyond" and "too far." The Enterprise of the Future understands the need to introduce innovation that the market is ready to accept and works to perfect its market-entry timing. It exploits the network effects of early adoption to take a commanding early lead.

The Enterprise of the Future *connects everyone to the customer*. Employees at all levels, from designers to warehouse employees, connect with customers through real-time information, online interaction, or, where possible, in person. The Enterprise of the Future also develops deep relationships with leading-edge customers and employees — those early adopters who determine market success or failure. It test-markets in these communities and collaborates with them to develop products. In the business-to-business space, the Enterprise of the Future invests to integrate its

systems with those of its key customers. This integration enables it to be a more proactive partner and an integral part of its customers' businesses.

The Enterprise of the Future *uses technology to anticipate shifts faster than the competition*. Market insights are critical to the Enterprise of the Future. It recognizes the value of the information it collects through its many channels and actively mines that information for insights. It uses emerging technologies, such as virtual worlds, to gain insights in new ways. It also puts systems in place that allow very fast feedback cycles. When customer preferences and demand start to shift, it knows before the competition.

Technology Strategy Implications

For CIOs, the main focus over the past 10 to 15 years has been business applications, with the goal of automating business processes to reduce cycle time and cut costs. To address these automation requirements, businesses have invested significantly in enterprise-wide applications such as enterprise resource planning (ERP), customer relationship management (CRM), and supply chain management (SCM). However, similar investments made by the competition have neutralized the competitive advantage of these application strategies. For most companies today, information managed by these systems is isolated, unsynchronized, and of questionable quality.

Information On Demand is the IBM approach for delivering the right information, at the right time, in the right context to drive business responsiveness and innovation. This initiative helps organizations unlock the business value of information for competitive advantage, providing business optimization and better business outcomes (Figure 1.2).

Figure 1.2: Delivering trusted business information with Information On Demand

Information On Demand requires a common information framework that defines common terms and models to describe the way information within a business relates across the business systems. This solution integrates, reconciles, manages, and analyzes data and content regardless of type, volume, or complexity. The sources can be from different silos or different vendors' systems and can be delivered in context, as quickly as required, to users, applications, and business processes.

The Need for Trusted Information

Today's CIOs have an opportunity to harness and leverage the information they've captured to provide sustainable competitive advantage rather than simply support or automate business processes. Yet accessing and verifying that information is not simple. For example, data is often locked up in multiple departmental silos, putting it at risk of duplication and other errors. This situation can also result in business units using different terminology or formats to describe the same information.

IBM clients have told us that for information to be considered "trusted," they need it to be accurate, complete, in context, and insightful (Figure 1.3). Achieving these elusive results across the enterprise requires a different kind of information infrastructure — one that allows information to be integrated and managed holistically.

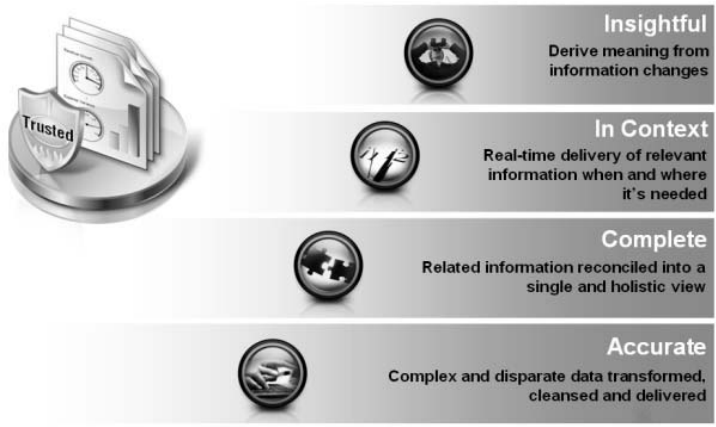

Figure 1.3: The importance of trusted information to business optimization

What is needed is a paradigm shift from siloed organizational information fiefdoms to the design and deployment of an enterprise-wide information infrastructure. Without such a structure, companies will continue to struggle in reaching the ultimate goal of optimizing business through advanced use of information.

As market pressures increase, organizations are looking for ways to harness information as a strategic asset. When treated as such, information becomes a significant competitive differentiator. All too often, however, this information is inconsistent, out-of-date, incomplete, or simply not available when and where it is needed most.

The effects of such information misalignment can be felt throughout the organization in rising costs, inefficiencies, and missed opportunities. A lack of trusted information can result in even more dire consequences. New products or business models executed on the basis of faulty information can result in costly failures.

Here are 10 common signs of an unstable information infrastructure:

1. No single enterprise view of data
2. Senior management requests for information that require intensive manual effort to respond to and far more time than desired
3. Low return on technology investments and higher operational costs
4. Multiple databases or spreadsheets storing similar data; no common data "dictionary" across the enterprise
5. No ownership of data
6. Difficulty complying with regulatory requirements, such as the Basel II Accord
7. Senior management questioning the quality, timeliness, or reliability of information used to make multi-million-dollar decisions
8. Difficulty answering questions about the origins of and business processes performed against data
9. Inability to keep up with the volume, pace, and variety of data
10. Need for extensive manual effort to understand source information

Organizations that make the creation and delivery of trusted information a priority can be richly rewarded. The potential return on investments includes numerous benefits:

- Extended protection for existing investments, applications, middleware, systems, and skills
- Enhanced visibility, control, and insight that enables organizations to optimize their business processes more effectively
- Improved data governance and management of data availability, consistency, integrity, and security in a way that meets both regulatory and management requirements
- Extended efficiencies from using information as a set of trusted, reusable services in a Services Oriented Architecture (SOA) environment
- Enhanced ability to integrate and use unstructured data, such as documents, customer feedback, and syndicated information from third parties

A New Approach: The Information Agenda

The information challenges are not new, but, as the CEOs we spoke with can attest, the need to address the problem is more pressing than ever before. Equally important, the ability to overcome these challenges and make timely, trusted information available has never been better.

For many years, the focus of IT transformation has been around the *application agenda*, driven by the need to increase automation, improve processing speed, and reduce costs. Although the application agenda remains very relevant in continuing to drive IT efficiency, many leaders are now turning their attention to the fact that their companies also need an *information agenda* to help drive business optimization and generate competitive advantage.

An information agenda is an approach for transforming information into a trusted strategic asset that the organization can rapidly leverage across applications, processes, and decisions for sustained competitive advantage. The information agenda is built on a solid approach that integrates strategy, information governance, and enterprise information infrastructure with inclusive implementation roadmaps. It is a comprehensive, enterprise-wide plan that the CIO, working with line-of-business colleagues, can create and implement to achieve strategic initiatives of the business. The goal is to transform the organization and unlock the business value of information across data and content silos.

Figure 1.4 depicts the four components of the information agenda, which organizations can use to deliver trusted, dynamic, enterprise information to optimize business performance:

- Information strategy
- Roadmaps
- Information infrastructure
- Information governance

Figure 1.4: The information agenda framework

Information Strategy

The information strategy establishes the principles that will guide the organization's efforts to create and exploit information to support its goals, objectives, and initiatives. It provides an end-to-end vision for all components of the information agenda and is driven by the organization's business strategy and operating framework.

As part of the strategy exercise, the organization should identify its corporate objectives and initiatives for innovation and differentiation. The corporate goals, objectives, and initiatives are the vision and strategy identified by senior management that outline the direction for the organization to maximize its inherent strengths and opportunities versus competitors. The information strategy should be revisited regularly to ensure it stays aligned with the organization's business strategy.

As an example, the diagram shown in Figure 1.5 identifies business initiatives and objectives strategic to an organization in the telecommunications industry. The information strategy must produce initiatives to support the identified business objectives.

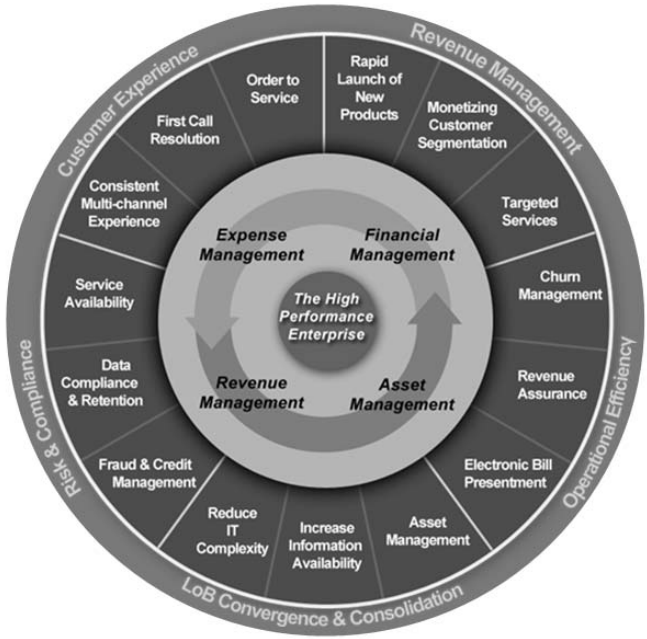

Figure 1.5: Sample business-driven information strategy

Roadmaps

The information roadmaps identify the sequence of iterative projects and objectives that return tangible results. The roadmaps can be effective in enabling the CIO to set expectations and to identify the sources of any needed funding. They should address

both short-term and long-term opportunities as well as a process for evaluating and tracking business benefits.

For the roadmaps to be most effective, the CIO needs to show where the organization stands now in terms of the maturity of its information use as well as where it wants to go in the future. An information agenda roadmap:

- Identifies and prioritizes projects that deliver a significant return on investment
- Identifies what data and content is important to the organization
- Identifies how and when the data should be made available
- Determines what capabilities are required to support and access relevant information
- Determines the management processes and governance practices that are required to sustain the information strategy over time

Information Infrastructure

Building an information infrastructure is the process of applying best practices, expertise, and advanced technology to achieve business optimization. A key step in this optimization is for the users to understand the data, either though business intelligence queries or business performance management applications. These applications cover the range from financial planning tools to dashboards to performance monitors. However, delivering this information on demand requires an infrastructure to transform the operational data into integrated data, organize the integrated data into trusted enterprise data, and deliver the trusted enterprise data as dynamic data to decision-makers across the organization.

To manage information as a strategic asset over time, companies must commit to an enterprise-level information infrastructure. Narrow, application-specific architectures result in significant operational inefficiencies, unacceptable turnaround on projects, and the continued proliferation of multiple copies of data and content. Within the context of an information agenda, an enterprise information infrastructure framework identifies the technology required to integrate current investments with future technologies, helping to optimize return on investment.

An integrated enterprise information infrastructure should have the following elements:

- *Business intelligence* to engage each and every decision-maker in an organization with information when, where, and how they need it; to make certain they are more informed about how to drive the business forward; and to ensure their decisions are more aligned with each other across departments — and to do so with little user training and minimal IT assistance.
- *Performance management* to enable business managers and executives to monitor how operational decisions tie to the overall strategic goals of the

organization and to apply this understanding to create accountable actions with forward-looking plans, budgets, and forecasts.

- *Data warehousing* to deliver the warehousing capabilities required to help organizations turn traditionally latent and historical data into relevant, real-time predictive analytics that enable timelier, more insightful business decisions.
- *Information integration* to accelerate the delivery of trusted information by enabling effective integration of information management tools such as enterprise intelligence (data warehouse) solutions, ERP and CRM deployments, and infrastructure consolidations (e.g., SAP).
- *Master data management* to enable master data — such as customer, supplier, partner, product, materials, and employee data — to be consolidated to provide an integrated, appropriate view of relevant, trusted information within any given business process.
- *Metadata management* to create and manage an organization's metadata, or information about information, and to define the meaning of data within a central, accessible repository. Consistency, completeness, and context of data are ensured via service, data, and content directories and translation, retrieval, and navigation processes.
- *Industry models* to enable the CIO and the organization to leverage the accrued knowledge and best practices of the industry to implement information integration, master data management, and next-generation data warehousing more rapidly and more successfully, with less risk.
- *Blueprints and analytic applications* to provide the functional and industry key metrics that organizations must understand, monitor, and adjust to accelerate the time-to-value of their trusted information.
- *Enterprise content management* provides content management, discovery, and business process management to support content-based business.

An integrated information infrastructure gives organizations the framework to address demands for real-time data access to trusted information, enabling optimization of the business through the advanced use of information (Figure 1.6).

Information Governance

Once you have your objectives aligned, roadmaps defined, and your enterprise information infrastructure in place, a key step of the information agenda is necessary to sustain the value of information over time. Information, like any other corporate asset, needs to be managed. This management — which we call i*nformation governance* — can enhance the quality, availability, and integrity of a company's information by fostering cross-organizational collaboration and policymaking.

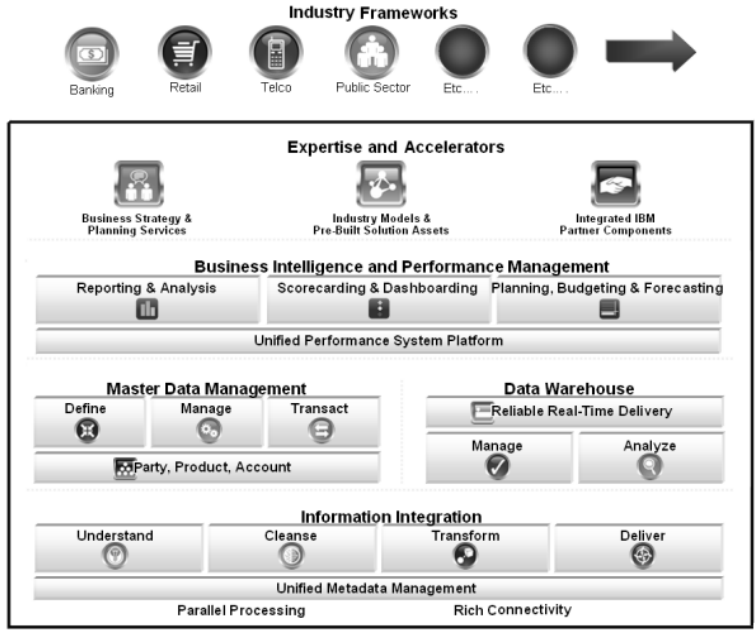

Figure 1.6: Integrated information infrastructure

Information governance requires the establishment of a specific corporate organization whose mission is to define the policies and practices for managing and protecting critical information assets over their lifecycle. The core objectives of information governance include the following:

- Defining governance infrastructure and technology
- Establishing common and standard information domain definitions
- Defining ongoing governance processes
- Developing architecture practices and standards
- Monitoring and ongoing improvement of data quality
- Identifying and training key business sponsors, governors, stewards, and executives of high-priority business areas
- Establishing the necessary organizational policies and cross-organizational oversight
- Training and enabling all affected staff

Although it is sometimes challenging for the CIO and line-of-business managers to work together to establish information governance standards, the importance of this part of the information agenda should not be underestimated. Robust information governance is necessary for an organization to comply with external regulations more quickly and more completely. Information governance can help unlock the fi-

nancial advantages that are driven by improved data quality, management processes, and accountability. Business performance can also be improved as a result of information governance — via common definitions and processes that drive effective strategy development, execution, tracking, and management.

Turning Vision into Reality

CIOs seeking to develop and implement an effective, business-aligned information agenda for the delivery of trusted information are going to need the assistance of their line-of-business colleagues. It is important, therefore, that CIOs obtain an executive sponsor — such as the board of directors, the CEO, or the CFO — that can position the information agenda as a strategic corporate initiative. This endorsement will help the CIO enlist line-of-business and other colleagues who are needed to participate in the development process and whose support will be critical to the success of implementation.

With strong executive sponsorship and a team of committed colleagues, CIOs can begin the accelerated path to an information agenda. To build its plan, the organization will need to go through several stages:

1. Vision lock
2. Business strategy and drivers review
3. Baseline assessment for maturity and capabilities
4. Future state development, opportunity identification, and prioritization
5. Integrated roadmap development and implementation planning
6. Kick-start project

These stages provide an accelerated, structured, and modular approach to defining, enhancing, justifying, and executing the information agenda. The stages are designed to align business and IT facets that impact the information agenda and to provide specific deliverables, including the following:

- Industry case studies
- Business value assessment
- Architecture readiness assessment
- Solution architecture
- Integrated project roadmap
- Software and services assets

Stage 1: Vision Lock

The objective of this stage is to create an agreement with executive stakeholders about the need for an information agenda and commitment to the following guiding principles:

- Trusted information is needed to make the decisions that deliver business strategy.

- Business functions should be responsible for defining the information needed to support their processes.
- Information should be efficiently provided as a shared service to all parts of the business.
- There should be one version of the truth for critical business information.
- Information should have one standard definition and presentation, unless compelling business differences dictate otherwise.
- Data quality and compliance using common standards should be built in at the source, with minimum intervention in the flow to the user.
- Information is a corporate asset. It should be freely shared within the business unless cost or legal/commercial sensitivity prevent it.

Stage 2: Business Strategy and Drivers Review

In this stage, participants identify the current business, organizational, and process models; define key customers; and determine the types of information needed to support these models and customers.

Stage 3: Baseline Assessment for Maturity and Capabilities

The baseline assessment defines the current state of the people, processes, and technology that make up the information agenda. The assessment includes:

- Readiness assessment
- Information maturity assessment and supporting data
- Data governance maturity assessment and supporting data
- People, process, and technology needed for existing and planned initiatives based on the stage 2 objectives.

Stage 4: Future State Development, Opportunity Identification, and Prioritization

For this task, IT leadership works with business leaders to agree on the desired future state of the information infrastructure and the key projects required to get there. This stage also includes prioritization of IT projects against corporate, line-of-business, and IT objectives to identify the first project to address with the information agenda. This stage includes the following elements:

- Target information maturity level
- Target data governance maturity level
- List of projects to accomplish target information and data governance maturity
- Identification of first area of focus — the “kick-start” project that can produce a “quick win” for the information agenda

Stage 5: Integrated Roadmap Development and Implementation Planning

Building on the findings of the first four stages, this step delineates the gaps in the organization's current information state and creates a phased execution plan to transform the organization and move to the targeted information state.

Stage 6: Kick-Start Project

While an all-encompassing, long-range plan is essential for transforming an organization, it can also be daunting. That is why IBM recommends identifying a specific "first project" and incorporating its planning into the overall approach to defining an information agenda. Specifically, this stage of the process creates a detailed execution plan for this first project, which was identified in stage 4. The execution plan includes:

- Agreement about key performance indicators to measure the success of the first project
- Selection and acquisition of technology components required for the project
- Incorporation of the appropriate information policies and procedures based on project requirements

The ultimate outcome of this process is the organization's information agenda. By bringing all the components into a single framework, the information agenda helps enable a closed-loop process for regularly re-examining objectives, strategies, and results to ensure they remain aligned with the organization's business strategy. Given the current business and technology climate of rapid change, this step is particularly crucial to long-range success.

2

Building Trusted Information for the Enterprise

A core element of the IBM information agenda is to help organizations better understand and leverage their existing information infrastructure. For example, if an organization wants to adopt a business performance management initiative to improve decision-making, it is better off discovering and then leveraging information from existing sources rather than going through the cost of collecting new information. However, in many cases, leveraging existing information "as is" is not an option, due to difficulties in finding relevant information and uncertainty as to whether the information, once found, can be trusted. Organizations need an approach to finding relevant information, transforming and integrating it to meet the needs of business performance management, and then delivering it to the necessary tools. Managing the information through this set of steps delivers the trusted information that forms the basis for relying on the results of business performance management analyses.

Finding, gathering, and integrating information is a critical aspect of business performance management initiatives. The value of information depends, first and foremost, on whether or not business users trust it. Simply collecting data without understanding its meaning or level of accuracy will result in poor information and therefore in poor decisions. Organizations commonly have at least some degree of inconsistent, incomplete, and incorrect information across their existing repositories. Many projects have failed or have been delayed significantly because data was simply "thrown over the wall" into a warehouse without appropriately addressing these issues.

Organizations can achieve trusted information through a systematic approach to information integration that goes beyond copying data to include aspects to understand the information and resolve inconsistencies. Master data management (MDM) builds on the information integration foundation and establishes trusted information for the core master data entities, ensuring the consistency of the most important business entities across business performance management applications and the enterprise. Enterprise information models address different levels of information

modeling with a goal of standardization across multiple competing perspectives, such as line of business or channel.

This chapter begins with a discussion of the typical business challenges that drive architecture and technology decisions. In subsequent sections, we discuss the role that information integration, master data management, and industry enterprise information models play in addressing these challenges.

Business Challenges

Historically, businesses have often underestimated the complexity of building integrated and trusted information. Most business performance management and warehousing initiatives that oversimplified or neglected information integration and equated it with "just copying data" have failed. To guarantee a positive return on investment of the information integration solution, it is important to consider the typical challenges that arise in this context.

Aligning Business and IT

The first, and most important challenge, is ensuring that the requirements and expectations from the business side are aligned with the design and implementation on the IT side. Business and IT users often speak different "languages" and lack the organizational structures, methodology, and technology to address these differences through improved communication and collaboration.

Unfortunately, it is all too common for business and IT experts to interpret the same business term in different ways. The same thing happens between various lines of business. For example, it is relatively easy to get multiple lines of business to agree that they need a report that includes customer information. What is considerably more difficult is gaining agreement across these groups about which data elements are relevant for the business concept of a customer. These discussions usually become embroiled in debates over naming, business terminology, and business rules. For instance, in an insurance company, a life-insurance line of business might consider it critical to distinguish between smokers and non-smokers, while a property-insurance line of business would not.

Worse still, these conversations sometimes never take place at all because a lack of precision in business terms hides the fact that disagreement exists in the first place. For example, what "customer" means may seem so obvious that the term is never clearly defined even though different types of customers exist (e.g., prospects versus important clients). When the meaning of critical business terms, such as elements in a report, are not clearly defined by the business, the IT teams may not realize the ambiguity and may make inappropriate architecture decisions based on their assumptions.

These considerations illustrate the need for a *common business vocabulary* that defines critical business terms and makes them easily accessible to business and IT

users. Defining such a glossary is core to the specification of enterprise architecture, and it is not a trivial task. We discuss the role of a glossary later, under the topic of enterprise information models.

Organizations cannot really address the challenge of a common vocabulary simply by defining business terms without ensuring that the vocabulary is commonly accepted across business and IT. A common vocabulary also must be managed to avoid unapproved changes that alter meanings unintentionally. For example, one line of business cannot simply declare that a "prospect" is equivalent to a "customer." Effective *stewardship* is therefore a critical success factor in developing a common vocabulary. Such stewardship ensures that the appropriate organizational structures and processes are in place to coordinate the definition of terms and control its accuracy.

Another challenge is how to achieve *ubiquity of the common business vocabulary*. The business vocabulary provides no value unless it is easily and securely accessible to all affected users — in particular, across business and IT. That means the vocabulary must be accessible in the normal work and tool environment of business users as well as IT personnel.

Another area that requires tight collaboration between business and IT is the identification of repositories and databases that contain relevant information. Once these data sources have been found, business analysts identify the relevant elements in them and, on a high level, identify how they map to the new integrated structures of the business performance management/warehouse solution. Business users need better tooling support than simplistic spreadsheets to specify these mappings in a format that is not too technical but lets them collaborate with the IT community. Although the use of spreadsheets for this task is common, it is also problematic. Because the spreadsheets are not linked to the actual underlying systems, elements are often missing or incorrectly specified. Even if the information is correct, copying it into data integration tools is frequently a manual and labor-intensive task. To sum up, organizations need a tool to support the business-driven specification of mappings and the generation of technical artifacts out of these specifications.

Understanding Your Information and Its Quality and Taking Action

A second category of challenges arises due to the fact that the information in existing systems is usually not correct, not complete, and not consistent across the enterprise, even if it is assumed to be. Information cannot be successfully connected if it doesn't meet fundamental requirements such as alignment of keys, availability of critical information, and so on. Data quality issues, especially if they go unidentified, pose a serious risk to the information agenda. At a minimum, they impact the value of the information being delivered because it will be not correct or complete, but out-of-date and not in the expected context. In the worst case, the information might be so poor that any use of it is impossible, resulting in project failure.

The process of understanding your information should start first by confirming the understanding of the meaning of critical terms. If it is not clear what "customer"

really means, it is impossible to assess the information related to this term. You have learned about the need to establish and maintain a common business vocabulary — an extremely important prerequisite to understanding and analyzing information and its quality. Based on this business glossary, we must then derive a logical perspective of the business concepts, their properties, and how they relate to each other. Ideally, this logical model extends to a full understanding of the business processes and other consumers that rely on this information. This model then provides a detailed understanding of the contexts within which information is used and further enables us to converge on reusable definitions of information structures across the enterprise.

After we have a sufficient understanding of the meaning of the information through a common business vocabulary and structurally through information models, we need to turn to the data itself and analyze it. Does the data really conform to the semantic and structural definition that is defined? Do any integrity constraint violations exist? What problems should we expect when integrating the data due to poor quality? For example, one automotive manufacturer started a project to integrate revenue information across previously separated divisions. It was assumed that the quality of information was acceptable, so the information was copied into the warehouse without any investigation. The first reports run on the newly integrated information showed zero dollars in revenue — an obviously surprising result. After some investigation, the architects realized that one system reported revenue in positive numbers, the other in negative numbers. The interesting aspect of this scenario is that problems in the integrated information were easy to identify because the numbers just didn't make any sense. But what if some of the numbers had actually been missing? This quality issue would not have been recognized, and business decisions might have been made based on wrong information. The requirement to address such issues is commonly referred to as the need for *data profiling*. Data profiling is defined as a specific technique or process, usually automated, to statistically measure and infer data content and relationships based on the actual underlying data as well as validate the data against technical and business rules.

The next step is to address any data quality problems identified through data profiling that negatively impact trust in the information. Many organizations acknowledge that their current information has data quality issues. Unless the business addresses these issues, business performance management applications will not provide valuable information. Consider the example of a company that wants to analyze customer profitability across multiple, currently isolated product lines. If the customer data (e.g., address information) is not cleansed and standardized, duplicate and incorrect information will negatively impact the result. If the same customer has two different addresses in two product-line databases, this customer cannot be identified as the same customer, and the analysis will lead to incorrect results — namely, the assumption that no joint customers exist across product lines. This requirement for *data cleansing* includes the standardization of data (e.g., to establish addresses with the same format), the enrichment of information (e.g., through adding missing postal codes in addresses), and the removal of duplicates.

To fully understand the information in a particular database that has been identified as a source for building the warehouse, it is often necessary to understand the context of the information and, in particular, where the information came from. Is this database the system where the information originates (e.g., through data entry), or has this system been populated through another data source? The identification of the ultimate origin of the information, its path from one database to another, and the type of operations between them strongly influence the trust we have in the information. Often, the same entity exists in multiple systems with inconsistent information; for example, the same customer has different addresses depending on which system is accessed. *Data lineage analysis* helps identify where information comes from and determine whether it can be trusted or needs to be further validated, cleansed, or transformed.

Integrating and Transforming Information

In most business performance management and warehousing initiatives, the information is not already available as required by the business. Instead, it is spread over multiple heterogeneous systems, in different formats, and often even with conflicting values. Even where data quality is not an issue, the conflicting requirements of operational and analytical environments necessitate the extraction of business information into a form that is suitable for business performance management applications. One of the most fundamental challenges of information integration is how to bridge this gap. How can we use the existing information and transform and integrate it so it can be delivered to the business users in the form they need?

Legacy systems often were designed and introduced many years ago, when the requirements from the business users were quite different from today. These systems continue to play an important role because they still manage some of the most important information assets in an enterprise. Often, they cannot simply be replaced in a short time just because business users need a new way to look at information or want to combine information differently. We need *data transformation* and *data cleansing* capabilities to extract information from information sources; transform, aggregate, and cleanse it; and load it into a warehouse for consumption by business applications.

Complicating the task of data transformation are the potentially very large volumes of data (e.g., all the financial information from a large worldwide company) and the need to process the information in a relatively short period of time. Extracting information from existing operational systems often negatively affects performance, so the data must be transferred during relatively short and continuously shrinking maintenance windows. After the data is extracted from the sources and while it is transformed, data cleansing rules must be applied to address data quality concerns.

Historically, information integration has been limited to batch processing during maintenance windows (e.g., on weekends). The frequency of performing information

integration tasks limits the currency of the information that business users access through business intelligence (BI) applications. Today's business users need to make decisions and take actions much faster, so, more and more often, a need exists for *real-time* data transformation and integration. However, not all use cases require real-time data integration. The technology must support different latency mechanisms for different tasks.

In some cases, it is neither possible (e.g., due to business or federal policies) nor feasible to integrate the information by copying it into a common repository. The only integration option is *virtual and on-demand information integration*. In this scenario, only when it is requested is the information pulled from existing sources, integrated, and then returned, without being stored along the way.

Delivering Trusted Information

The primary channel discussed in this book for delivering information is through a business intelligence or business performance management framework. As noted above, business users who interact with such applications need access to information that often resides in disparate sources and requires integration and transformation. One of the remaining challenges is how to ensure that these applications do not establish yet another silo. Many other applications and business processes also need access to trusted information. How do we prevent each application from accessing, integrating, and transforming the data in different ways? How do we ensure that future and strategic information delivery channels do not yet again become inconsistent?

The consistency, correctness, and completeness of information is especially challenging for one of the most important information assets in an organization: *master data*. Due to its importance, master data is ubiquitous, existing in almost all systems. But because existing systems are often isolated and lack the proper integrity controls, master data is also very inconsistent, incorrect, and incomplete. Master data plays a central role in information delivery. Many applications and processes need to access this data. The key challenge here is how to provide consistent, correct, and complete master data given the limitations that the existing systems cannot simply be replaced and that so many different applications — analytical, transactional, collaborative, and so on — need to consume this information.

Certainly, the delivery of consistent, correct, and complete information is in particularly important for master data. However, it is important for any data, whether master data or any other kind (e.g., unstructured information such as loan applications, financial data that includes key performance indicators). The challenge is how to guarantee *consistent delivery of trusted information* across a broad range of information consumers and for any information.

Maintaining Trusted Information

Organizations sometimes treat information integration projects as a task with a completion date defined as the point in time when all the information is aggregated, transformed, and available in a warehouse. At that point, the major task is considered to be accomplished, and the remaining work is just to refresh the warehouse from time to time. This approach would be appropriate if nothing changed — if source data were never altered in a way that could impact the meaning, quality, or context of the information. Unfortunately, this assumption is usually an unrealistic one. The challenge in today's dynamically changing environment is how to maintain trusted information over time.

The more valuable information is (e.g., due to the decisions made with it, the number of consumers of it), the more important it is to ensure that the trust of that information does not degrade over time. Often, organizations do not properly control the entry of new information or requests to modify information. Sometimes, the integrity of information to be updated is not validated at all or is validated insufficiently or inconsistently across different applications. The challenge is how to maintain the same rules to ensure the trust in information over time. Which organizational structures, processes, and technologies are needed to address this challenge?

Methodology for Building Trusted Information

We have discussed some of the typical business challenges related to information integration. When an organization tries to address these challenges, more than just one task is often necessary. The remaining challenge, then, is determining the best overall approach to address this complicated problem. What should be done first, what needs to follow, and in what order? What parameters control the flow of these activities? Is there a systematic approach to this problem space? In other words, is there a *methodology for building trusted information*?

As you will see in the following sections, organizations can use information integration, master data management capabilities, and the leveraging of enterprise information models to address these major challenges.

Information Integration

Many organizations have started to realize the need to better leverage their information in order to make better decisions and reduce the time people spend trying to find information. Even if users find the information they're looking for, they often must manually transform it and assemble it in their head rather than being supported by an IT system. What is worse, they don't know whether they can trust the information they find.

Information integration is one of the key components in building trusted information. The main objectives of information integration are to identify the (disparate) sources that provide the required information; to understand those sources, the

information in them, and its characteristics; to cleanse, transform, and merge the information; and to deliver it to consumers. We will introduce the major information integration capabilities by grouping them into categories and presenting them in a sequence that represents a high-level recommended methodology. After this overview, we'll describe some of the key functions in more detail.

- *Understand* — This set of functions can help companies to automatically discover, model, define, and govern information content and structure, as well as to understand and analyze the meaning, relationships, and lineage of information. This is the first set of tasks that needs to be performed jointly between business and IT. The alignment of business requirements with the understanding of the existing IT environment, as has been noted, is a key business challenge. Areas of business understanding encompassed by this set of functions (and described further later) include the business glossary, business-driven design and automation, and data profiling.

- *Cleanse* — Data cleansing supports the quality and consistency of information by standardizing, validating, matching, and merging data. This capability is a key enabler to improve the quality of information and to take action to address quality issues — another of our key business challenges. We will discuss the concept of data cleansing in more detail later in the chapter.

- *Transform* — Information transformation combines and restructures information to help ensure it is in the proper context for new uses, addressing the business challenge of integrating and transforming information. This is the core functionality from which information integration has initially started: extracting the information from its sources, transforming it, and loading it into a target database, such as a data warehouse or a master data management repository.

- *Deliver* — An increasing number of business scenarios require real-time integration and transformation. Information changes must be captured as soon as they occur and then replicated and/or published to other databases or directly to applications and processes. Another integration option is to provide a virtualized access over disparate sources so that information is pulled on demand from its current sources and integrated on the fly through data federation. The third function in this category is the consistent delivery of information across various applications and processes. All the concepts we have introduced so far guarantee such a consistent delivery in the traditional information processing context — for example, supporting business performance management initiatives. Enabling the delivery of trusted information as a service extends this reach to any consumer in a Service Oriented Architecture (SOA).

Unified Metadata Management and Deployment

An information integration approach needs to be built on a unified metadata infrastructure that enables shared understanding between the different user roles involved in a data integration project, including business, operational, and technical domains. This common, managed infrastructure helps reduce development time and provides a persistent record that can improve confidence in information while helping to eliminate manual coordination and synchronization efforts.

One of the key business challenges introduced in the beginning of this chapter was the alignment of business and IT. Multiple user groups with a range of different roles from the business side as well as the IT side must be able to seamlessly and directly share information with each other across organizational boundaries and different roles. Streamlining collaboration among multiple, disparate users is critical to the success of any integration effort.

The roles required in information integration projects typically fall into the following categories:

- *Project managers* have overall responsibility for the project, including training, deployment, management, resource allocation, coordination, and progress tracking.
- *Business analysts* provide in-depth business background and context to ensure that integration implementations meet business requirements and needs. *Subject matter experts* interpret end-user business requirements and assist the project team in prioritizing requests, defining business terms for context, documenting business transformation logic, and developing the project validation criteria.
- *Data stewards* have deep business and technical knowledge that often bridges the gap between business and IT. Stewards manage logical data resources and coordinate data definitions, aliases, quality control, improvement efforts, access authorization, and planning for subject-area data.
- *Architects* help ensure that enterprise standards are met and applied consistently across the enterprise and across projects. *Information architects* often manage the process to standardize efforts related to metadata creation, maintenance, enhancement, and distribution. They also work with developers to advocate data integration best practices in the development teams.
- *Integration developers* create processes and jobs to manage data manipulation. For example, *data cleansing developers* design and develop complex data cleansing applications. *Data integration developers* create and test "extract, transform, and load" (ETL) applications to support data integration and cleansing solutions. *SOA developers* deploy data services and ensure that existing services are reusable and will meet service level agreements (SLAs). SOA developers also create service documentation for access by other members and maintain services and registry information.

- *Administrators* oversee key areas of system implementation. *Database administrators* (DBAs) are in charge of database commissioning, installation, configuration, deployment, and management as well as data modeling. *System and software administrators* are responsible for project hardware, software, configuration, environment maintenance, and deployment as well as management of system users and their associated security roles.

Information integration projects are often part of a larger initiative, such as supporting a business performance management initiative, building a master data management solution, implementing a Service Oriented Architecture, conducting application migration or consolidation, or other enterprise-wide deployments. The roles listed above represent the subset that is required from an information integration perspective. They will need to participate in the broader initiative and collaborate accordingly (e.g., with enterprise architects, security architects).

IBM InfoSphere Information Server

The *IBM InfoSphere Information Server* provides a platform that realizes the information integration capabilities described above. It includes a unified set of product modules designed to streamline the process of building an information integration solution (Figure 2.1). It enables the various roles to collaborate effectively and execute the key functions of an information integration project — that is, to understand, cleanse, transform, and deliver trusted information. Organizations can use information validation, access, and business processing rules across multiple projects, leading to greater consistency, more control over data, and improved efficiencies.

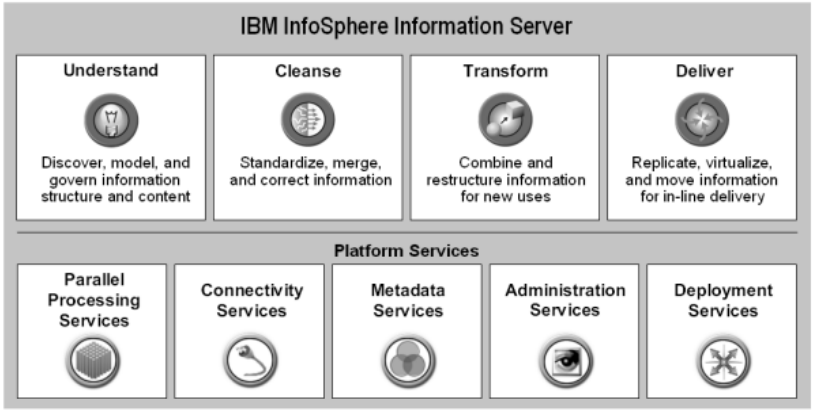

Figure 2.1: IBM InfoSphere Information Server's information integration capabilities

IBM's InfoSphere Information Server offers a set of services built on a common metadata management environment, enabling information to be shared seamlessly among project data integration tasks:

- *Parallel processing services* ensure that the information can be integrated and delivered with the highest degree of performance and scalability, at a minimal cost.
- *Connectivity services* enable the various tasks and functions to connect through the same mechanisms to a broad range of source systems with highly optimized access. It is essential to have the flexibility to connect to a broad range of source systems, but it is also important to connect to them most effectively.
- *Metadata services* provide a unified metadata management platform that supports the communication and collaboration between the various information integration tools, functions, and roles.
- *Administration services* permit the management of the various tasks in a common fashion and in a consistent environment.
- *Deployment services* assist in the deployment of information integration functions across a variety of hardware platforms with minimal configuration changes.

With this introduction to information integration as background, we will now examine some of the key capabilities in more detail.

Business Glossary: A Common and Controlled Business Vocabulary

The key challenge addressed by a business glossary is the alignment of business and IT, focusing specifically on a common vocabulary or "language." When business users talk about a customer, they may mean different things because they imagine what they could do with customer information in a new application or a new business process. IT users may have a different understanding because they may think about customer information in a different context, for example as it is currently stored in a particular system.

A business glossary, sometimes called a dictionary (or data dictionary), contains definitions of terms used by the organization to support business initiatives. The glossary defines the language of the enterprise and, by extension, the language of projects — providing the collaborative link between disparate groups involved in integration efforts. Without a formal glossary to capture and centrally manage this valuable company asset, organizations run the risk of this critical information leaving the building each day when employees go home. They also risk having the various roles across a project possessing a different understanding of which information to actually deliver. It is often assumed that the meaning of "customer" or "member" is clear when in fact it is not clear, for example, whether a prospect is a customer or not.

A glossary enables data analysts, business analysts, and subject-matter experts to create a rich glossary, linking business concepts to technical metadata and exposing these linkages across the enterprise through easy-to-use, simple interfaces. The business glossary facilitates the creation of comprehensive, authoritative terms and defined relationships via categorical hierarchies. Organizations can assign data stewards to manage this information to support data governance initiatives that require accountability and responsibility for topics, assets, and relationships.

A business glossary needs to support the roles and responsibilities of the data steward and subject-matter experts to create rich, detailed definitions for terms and to define categories to represent the relationships between the terms. In addition, business analysts should leverage a business glossary to link technical artifacts, such as database tables and columns, to business terms. This linking helps ensure that particular data artifacts are coupled with their business context, and it enables two-way communication. Business users can drill down from a term to find the technical data sources, while technical users working on a data source or ETL job can understand the business context of the objects being used.

Just as important as building and managing the glossary is the ability to expose its content to different users across the organization. An understanding of the business context enables organizations to create better applications to meet business requirements. A business glossary is primarily a tool for business users. It needs to be designed in a way that lets anyone in the organization easily view the contents of the common glossary, and it should promote the adoption of a standardized language across the enterprise. Business users are not familiar, not should they necessarily be, with technical tools. The easier it is to invoke a glossary directly from any application, the higher the adoption rate and success of the common vocabulary will be. Users should be able to search for any term without losing the context of the application they are currently using and should be able to easily access all relevant and associated information for a term, including a definition, examples, the term's steward, and so on.

However, as we noted above, a business glossary should not be limited just to the business audience. It must also be available to the IT side in their preferred environment and, therefore, tightly integrated with technical tools that analyze, cleanse, and transform the information.

The *IBM InfoSphere Business Glossary* provides such a common and controlled business vocabulary. Business Glossary uses a browser-based interface with business-user–oriented interfaces. Users can also invoke the glossary content through a simple text selection from any application to display relevant content in a lightweight window without losing the context of the primary application. The content in InfoSphere Business Glossary is exposed to other IBM InfoSphere Information Server modules to enhance the collaboration between the various business and IT roles.

InfoSphere Business Glossary enables the alignment of business and IT by commonly defining, controlling, and sharing the vocabulary of the information

integration project across all participating roles. InfoSphere Business Glossary significantly reduces risk by eliminating the ambiguity of business terms and requirements related to the information that needs to be delivered.

Business-Driven Design and Automation of Information Integration

As we have noted, one challenge when aligning business and IT and improving their collaboration is the common business vocabulary. A business glossary addresses this problem by establishing and controlling common terminology and mapping the business terms to technical artifacts. For example, a business glossary defines the meaning of "customer" and associates it to tables and/or columns in one or more databases. This association of business terms to columns, tables, and databases helps identify relevant sources when integrating information. For example, when we need to build a dashboard for "customer" information, the associated databases in a business glossary can provide a first and high-level glimpse of which sources we may have to integrate.

Business analysts also need to specify in more detail what information is required, which sources to retrieve it from, and whether transformations are required. Typically, business analysts and subject-matter experts are responsible for specifying *what* type of transformations are necessary. The IT side then takes these requirements and determines *how* to implement the transformations. Another challenge, then, is the collaboration between business and IT when transforming the business requirements into technical mapping implementations. Often, business users use spreadsheets and text documents to specify their requirements. Although these methods let them enter requirements quickly, they also introduce errors because the spreadsheets are disconnected from the real underlying systems. Business analysts can easily forget to include important elements, mistype them, or even include items that do not exist. Architects and developers often manually copy and paste these requirements into their IT tools, which is the next source of errors.

A tool that implements business-driven design and automation of information integration tries to reduce the inordinate amount of development time that organizations today waste in overcoming differences in languages, skills, and working methods; in clarifying business requirements; and in synchronizing tool output. The goal is to break down these barriers to help maximize team collaboration, increase automation, and ensure projects are completed successfully and on time. Companies can achieve these goals through an integrated environment that includes business analysts, developers, and data modelers to accelerate collaborative development across user roles, products, and geographies.

Such a tool automates efforts across multiple data integration tasks while incorporating business perspective and maintaining lineage and documented requirements. It gives business analysts and subject-matter experts an interface that is as easy to use as a spreadsheet. At the same time, it is connected to the underlying existing systems so that business users can easily access important metadata — that is, characteristics

about the information, such as which tables exists, what their attributes are, what glossary terms are associated, and so on. This capability significantly reduces errors in developing the mapping requirements and initial business specification.

Where complex transformations are required, the tool lets business users enter mappings as text or annotations. For simple mapping requirements — filtering certain values (e.g., '999999'), for example — the tool provides easy-to-use wizards that enable business users can enter technically correct specifications.

Integrated into the development platform of the architects and developers, the tool lets these users easily access mapping requirements, review them, and analyze them from a technical standpoint. In cases of simple transformations where business users have provided technically correct mapping specifications, the tool can automatically generate the technical artifacts (i.e., a data flow job) that the technical users can then deploy, reducing development time. Even in more complicated cases, the tool generates a skeleton of a deployable data transformation job and adds the annotations from the business users, giving developers a jumpstart on completing the technical mapping.

IBM InfoSphere FastTrack offers a tool for business-driven design and automation of information integration as described above. Part of InfoSphere Information Server, InfoSphere FastTrack is integrated with the other modules from a tooling as well as a metadata management perspective. For enterprises of all sizes, FastTrack speeds time-to-value for strategic initiatives by accelerating the translation of business requirements into data integration projects that solve complex business problems.

Data Profiling: Understanding the Quality of Information

Over time, data stored in legacy systems and enterprise applications can lose much of its value as data characteristics and the general knowledge about the data changes. Although the data may have served its original purpose perfectly, it is often inappropriate for use in other applications as time passes and requirements evolve. And often, the actual data does not conform to its initial definition and integrity constraints. For example, although the customer identifier is supposed to be a number, in reality it may include faulty characters, or some identifiers might even be missing. The larger the scope of the data and the more source systems involved, the more impossible it becomes to identify these issues manually. If the issues are undetected, they can significantly complicate the information integration project: records across systems cannot be joined because their identifiers have unknown differences in formats, source records cannot be loaded into a target because of data type violations, and so on.

Data profiling is a technique that helps improve the understanding of the actual data that is stored in the various source repositories and needs to be integrated and transformed. Data profiling examines the data values, inferring additional characteristics from those values and then assessing structural integrity and relationships across data sources.

Data profiling starts by capturing existing metadata — characteristics about the information such as definitions of schemas, tables, files, and columns. Column analysis — for example, investigating a column such as "postal code" — looks explicitly at distinct values, enabling the tool to generate inferred metadata and determine the true physical characteristics of the data. Key and cross-domain analysis processes help highlight duplicated data, broken keys, and missing or invalid data relationships across tables. Such problems can affect business processing, data migration efforts, or the loading of critical data into production systems or data warehouses.

Analysts working with the data profiling results can enrich them by creating annotations that can be shared across other tasks. For example, it is valuable to evaluate the differences between how the information is supposed to be structured and its actual format, to identify incomplete or invalid values and other issues.

Data profiling fulfills a critical role in the integration process. Profiling data helps the analyst completely understand the data sources before beginning to design the detailed mapping specifications. It can also help validate and refine mapping specifications that have been specified in a tool for business-driven design and automation, such as InfoSphere FastTrack. It can also guide the detailed design of cleansing and transformation operations that will need to be deployed as part of the information integration solution. The more information the analyst has about the true data structures and content, the more accurate the requirements are for the downstream developers. And the earlier that inconsistencies and data issues are surfaced, the lower the costs are to address them. Problems that are not identified until the information integration project is deployed and information is being delivered can cause inaccurate decisions and expensive cleanup projects.

IBM InfoSphere Information Analyzer provides data profiling capabilities that can help reduce project costs and risk by discovering problems in the early stages and monitoring changes in data structure and content. This tool offers a comprehensive analytical view into the characteristics of a company's data assets through a repeatable, proven process. It is integrated with the other InfoSphere Information Server tools so that users of any of these tools have access to the data profiling results when providing an understanding of the data and when designing cleansing and transformation operations.

Data Cleansing: Improving the Quality of Information

Corporate information often resides in a variety of repositories (databases, files) with different scopes, contexts, and formats. Over time, the needs in organizations change due to business expansion or contraction, new business opportunities, and other changes. Correspondingly, the way an organization needs to process and deliver information changes. Driven by short-term return-on-investment constraints, organizations often tend to build yet another application and database that have their own unique ways of representing and structuring information. Evaluated by

themselves, each application and its repository meet, in many cases, the very specific needs in a particular context. However, when information needs to be gathered across the organization, not just pulled from one silo, significant problems emerge.

The information agenda, and specifically business performance management initiatives, supports critical executive decision-making that needs to be based on a broad range of information and must not be limited by IT system boundaries. For example, customer information often resides in a variety of systems, and many organizations lack a common and consistent identifier of customers that would permit a painless aggregation of this information. In most cases, the address information is not even standardized but has inconsistent representations of the same location. We have discussed the need to understand the quality of information. If the quality of the information doesn't meet the requirements of the business, technology must be applied to establish trusted information.

Data cleansing automates the process of investigating, standardizing, matching, and surviving data. The *investigation* task parses, classifies, and analyzes patterns in source data. *Standardizing* the data ensures that each data type has a correct and consistent content and format. Some of the standardization rules are straightforward and require little significant effort, such as converting mixed-character data into upper case. Other rules are relatively advanced and require access to a database that stores correct values, such as the correct association between postal code, city, and state in the United States. Standardization rules may also need to be context-sensitive. For example, a string such as "St. Virginia St." is determined to have a street name of "St. Virginia" and a street type of "street" (assuming this is a U.S. address). While technically the same, the two occurrences of "St." have different meanings when interpreted through the eyes of an intelligent rule set.

Unfortunately, even after standardizing, the data values in the records often are not identical. In one record, a person's name might be "J. Smith"; in another, it might be "John Smith". One challenge when *matching* records is to determine how likely it is that "J. Smith" is the same as "John Smith". Obviously, the answer depends on the other information contained in the record. If the address is exactly the same, the match might be very likely. The answer will also depend on how many people have the same first name and last name: it is less likely that two "April Back-Cunninghams" exist in the same city than two "Robert Johnsons." Although these examples have used mostly U.S. addresses, this approach certainly applies to any country and not to just addresses. In fact, the standardization capabilities of data cleansing tools should not be limited to certain countries or regions but should allow for worldwide use.

Survivorship and formatting ensure that the best available data survives and is correctly prepared for the target destination. If necessary, the existing information is enriched with data from external sources (e.g., missing postal codes).

This data cleansing process needs to be applied when consolidating large volumes of information from many sources in one batch — as when building historical

snapshots in an enterprise warehouse to support a business performance management initiative, for example. This step will ensure high-quality, accurate data and the consistent, correct identification of core business information throughout the enterprise. The technology must therefore permit cleansing information at high speed to reduce the time to load the warehouse. However, the cleansing technology should not be limited to applying cleansing to persisted data in a database that is moved from one place to another. It should also support a Service Oriented Architecture so that cleansing rules can be exposed across the enterprise and information can be validated during data entry before it is even stored.

IBM InfoSphere QualityStage offers a data cleansing technology with probabilistic data matching capabilities and dynamic weighting strategies. It creates and ensures high-quality, accurate data by consistently linking and consolidating core business information (e.g., customer, location, product) throughout the enterprise. This capability helps reduce the time and cost of implementing strategic, domain-focused projects by improving organizations' understanding of their data. The business rules defined to combine and consolidate records are stored in the metadata repository and shared with other modules of IBM InfoSphere Information Server for full insight and auditing into the cleansing and data consolidation processes.

Transforming and Consolidating Data

As we noted in the discussion of business challenges, information required by the business is most often not already available in the expected scope, form, and context. It must be gathered from the various sources and transformed into trusted information that can then be delivered to the business. Complicating this task are the ever-increasing volume of information, the complexity of operations to restructure that information into the appropriate format, and time constraints. Although the duration of transformation operations — particularly when processing large volumes of information — is not in the range of seconds but in hours or even days, how long such transformation processes take does matter. Business performance management–related decisions often depend on the availability of information within a relatively limited time period. Business users are continuously demanding shorter processing cycles to build integrated, trusted information.

Data transformation and consolidation consists of three major phases: the extraction of information from the sources, the transformation of the information, and the loading of the integrated information into a target database. Data transformation technologies need to support a broad range of sources, including all varieties of databases as well as packaged applications, files, and so on, on different hardware platforms. Performance and scalability is an important requirement throughout the complete transformation process, from extracting to loading the information. When measuring the degree of performance and scalability, the effort to tune the system and the tooling support is often a critical requirement. Can the data transformation process be specified independently of the underlying hardware and scalability

requirements, or must it be adjusted when additional hardware is added? Does the technology provide tools to estimate the required hardware resources to better project costs and plan the rollout, or is the administrator left on his or her own to proceed with trial and error? Last, how much tooling support does the technical team possess to analyze performance and possibly to improve it? Going manually through log files to determine the length of a transformation job run is time-consuming and does not provide the level of detail necessary to most effectively tune a system. The data transformation tool must also allow integrating the data quality operations into the overall transformation process and execute them either independently or in a tightly integrated way.

IBM InfoSphere DataStage leverages the parallel execution capabilities of IBM InfoSphere Information Server to meet companies' most-demanding data volume and transformation requirements. It has been certified by a third-party benchmark auditor accredited by the Transaction Processing Performance Council to scale almost linearly as hardware is added to the processing environment. In addition to running on Linux, Microsoft Windows, and UNIX platforms, InfoSphere DataStage runs natively on mainframe IBM S/390 and IBM System z Linux environments, enabling organizations to fully leverage their IT investments. It has the unique capability to allow architects to design a transformation process once and deploy it flexibly on different architectures ranging across uniprocessors, symmetric multiprocessor, and clustered, massive parallel processing, and GRID systems.

InfoSphere DataStage ETL jobs consist of technical design metadata that describes the job flow and transformation logic being applied. InfoSphere Information Server has the unique capability of linking the design and operational metadata together to provide a complete picture of what actually happens when a job is executed in a production environment at any point in time. This functionality is critical for supporting and troubleshooting complex integration environments, as well as for supporting compliance reporting for tracking data lineage.

Real-Time Data Integration and Change Data Capture

In today's fast-paced world, access to real-time data has never mattered more. To be successful, organizations must be able to report and analyze corporate data quickly and easily, regardless of what applications created the data, which platform they're running on, or what database they're stored in. Businesses need to synchronize inventory, financial, and customer information between existing systems and Web applications. And they need to be able to consolidate and distribute data in real time between applications across different regions, business units, and departments. Organizations often need to process data immediately, as soon as it changes and without delay.

Technologies that deliver transformed and integrated information in real time are based on the concept that information is processed as soon as it changes. Changes in the sources are captured as they occur, potentially transformed, and

then distributed to the target system, which can be a data warehouse or a standby system that functions as a replicated database for failover. To the data transformation tools, the only difference is that data is not extracted in batch volumes at relatively infrequent intervals; instead, any changed data — even single entries — is transformed continuously. It is therefore important for the change data capture technology and the data transformation technology to be based on a common framework so that information can be moved in batch and/or in real time using the same transformation logic.

Change data capture technologies can be applied to load data warehouses in real time so users can make operational and tactical business decisions based on the latest information. These capabilities are therefore a critical component of a dynamic warehousing initiative. Organizations can also use this technology to populate real-time dashboards for on-demand analytics and business performance management initiatives.

In addition to traditional batch processing, InfoSphere Information Server supports real-time data integration and replication with *IBM InfoSphere Change Data Capture*. The InfoSphere DataStage tool can source information from InfoSphere Change Data Capture using low-impact technology to provide scalable, high-performance, heterogeneous data integration without impacting the source systems. The option to leverage real-time data integration lets customers get the information they need, when they need it, to help them make decisions at the speed of business. With real-time data integration solution, today's organizations are making better business decisions, running smoother operations, winning new customers and partners, and increasing their bottom line.

Data Federation: On-Demand Access to Virtually Integrated Data

A third option for integrating and transforming information — in addition to batch transformation and change data capture — is to provide a virtualized access to distributed information. The two previous approaches create transformed information that is stored in some repository, such as a database or a warehouse, from which it can be accessed. But moving data to a centralized location is not always feasible. The volume of data might be very large, or the information might change very rapidly. The consumers of the information may need only relatively infrequent access to the integrated information, or they may require the most up-to date integrated information. Given these parameters, information integration based on data transformation or change data capture may result in relatively high overhead.

The *data federation* approach leaves the data in place and does not create a new copy when integrating it. Instead, when a consumer requests access to distributed information, the data federation technology integrates the information on the fly. It accesses the distributed sources only on demand according consumer requests, joins it together, and delivers it immediately to the requesting application.

Performance is an important consideration in most areas, but it is of particular significance when federating information. If an application sends a request for integrated information to the federation server, it must wait until that server has gathered the distributed data, integrated it, and returned it.

Data federation technology complements the other integration approaches. Business performance management initiatives can greatly benefit from this technology. Distributed and heterogeneous data can be accessed in place and made available in real time. Some organizations have started to apply these complementary integration options. They use the data transformation approach for information that requires complex transformations and needs to be accessed repetitively by many users with very high and predictable performance. They complement that solution with federation technology for portions of the data that are more volatile and require less frequent but real-time access.

IBM InfoSphere Federation Server integrates information across disparate data sources in real time, providing advanced query optimization techniques to minimize response times. Federated queries can be exposed directly within a data transformation job flow to help expand connectivity and simplify the transformation design flow.

Information Service Enablement: Delivery of Information Through SOA

Once we have established trusted information through some of the mechanisms introduced above, we need to ensure that this information is delivered in a controlled but also consistent manner across the enterprise. In this book, we focus primarily on business performance management and business intelligence scenarios. However, organizations must deliver information to a broad range of consumers, including applications, business processes, and portals. A Service Oriented Architecture provides the appropriate platform to publish information through consistent, reusable services to a broad range of consumers across a heterogeneous IT landscape.

Using *IBM InfoSphere Information Services Director*, architects can build information services that expose trusted information by exposing InfoSphere DataStage, InfoSphere QualityStage, InfoSphere Federation Server, and InfoSphere Classic Federation Server for z/OS and IBM DB2 logic as services. InfoSphere Information Services Director provides built-in load balancing capabilities to help ensure fault tolerance and high availability.

Lineage and Impact Analysis Through Metadata Management

Information integration projects that consider the initial delivery of trusted information as the final milestone are most often considered a failure in the final analysis. Even though the initial load of a warehouse with trusted information is a key milestone, it is important to understand that the IT environment will change at some point and that we need to be prepared for that event. If the data volume changes in the sources or nonfunctional requirements become more aggressive —

such as demanding faster delivery of information — the information integration technology needs to be able to accommodate and scale accordingly. This is typically accomplished with powerful integration engines such as InfoSphere Information Server. However, an information integration technology cannot always automatically adjust to changes to the existing systems. For example, if new columns or tables are added or the delivery format of trusted information changes, it is not always clear how to modify the affected transformation processes. What's worse, it often is not even clear what is affected by the change. Historically, businesses have addressed data transformation problems with proprietary code, such as FTP programs and scripts that are not documented at all. In this situation, it is difficult to identify the impact of a change and adjust accordingly.

To avoid this problem, we must capture how the information flows between sources (upstream systems) and targets such as a warehouse (downstream systems). When we build transformation processes, the information integration platform needs to capture the way the information is flowing and make it available for later reuse. This mapping specification can then be used for impact analysis. Based on what is about to change (in an upstream or downstream system) and which transformation processes are related to this scope, a tool can then highlight the impact of that change.

Capturing this information flow (i.e., the metadata) helps us to understand the information's complete *lineage*, including where data comes from, what it is related to, and what happened to the data as it moved across applications and data warehouses. This metadata is also valuable input to reports that support compliance and governance/regulatory standards such as the Sarbanes-Oxley Act and Basel II.

IBM InfoSphere Metadata Workbench promotes metadata reporting, management, and insight across the InfoSphere Information Server modules. It gives developers and administrators a Web-based interface to explore and navigate InfoSphere Information Server metadata and third-party metadata touch points to reporting and modeling products. Metadata Workbench enables users to explore key information assets and understand their use, relationships, and meaning. It provides a comprehensive view of metadata to help businesses understand the lineage of information from the upstream systems through the warehouse all the way up to the report and to analyze the impact of changes in that chain.

The search and query capabilities of InfoSphere Metadata Workbench are designed to enable developers, administrators, managers, and analysts to view, understand, and explore metadata across InfoSphere Information Server to modeling and BI tools. This unique visibility promotes understanding and re-use on new projects and ultimately can reduce duplicate development effort, shorten development times, and enhance efficiency.

Master Data Management

The management of key organizational information has always been important. Being able to identify the customers, products, and services offered by an organization and the arrangements or accounts that result with those customers and suppliers is fundamental to the operation of most organizations. Whether the organization is a bank, a retailer, or a government agency, a core set of such information is used across the enterprise. This information is *master data*.

Master data is some of the most valuable information a business owns. It represents core information about the business — such as customers, suppliers, products, and accounts — and the relationships between them. Each of these domains of master data represents information consumed across different business processes, across organizational units, and between operational and decision support systems. A common understanding of this information is useful, both to prevent bad things from inadvertently happening, such as posting a bill to the wrong address, and to provide an opportunity for significant business benefits, such as improving the ability to sell complementary products to customers.

Management of master data is not new. Most organizations have systems to store and retrieve the master data that is critical to their business. Unfortunately, many information systems have become increasingly complex in response to the pressures of growth, business changes, and technology developments. Multiple, often redundant, stores of master data can be found throughout many enterprises. It is not unusual, for example, to find more than a dozen different (and often inconsistent) repositories of customer data in an organization. It has therefore become increasingly difficult for organizations to identify, maintain, and use an authoritative set of master data in a consistent way across the enterprise.

When master data is spread across multiple systems in an unmanaged way, multiple competing views of master data may exist — different customer lists, lists of suppliers, or product definitions. Without a complete and authoritative set of master information, it is difficult for enterprises to optimize their relationships with customers and suppliers across product lines. It is hard to rapidly introduce new products to the market or relate sales performance to product categories.

In an ideal world, there would be a single place where all common master data in an organization is stored and managed. The data would be accurate, consistent, and maintained in a coherent and secure way. All updates would take place against this single copy of master data, and all the different users of master data would interact with this single authoritative source of information. The goal of master data management is to enable this ideal world. Through a combination of architecture, technology, and business processes, MDM provides an approach to incrementally reducing the amount of redundantly managed information and providing information consumers throughout an enterprise with authoritative master data.

It is useful to consider master data along three distinct dimensions: the *domains of master data* that are managed, the *methods* by which the system is to be

used, and the *styles of implementation* that are needed for a particular deployment (Figure 2.2).

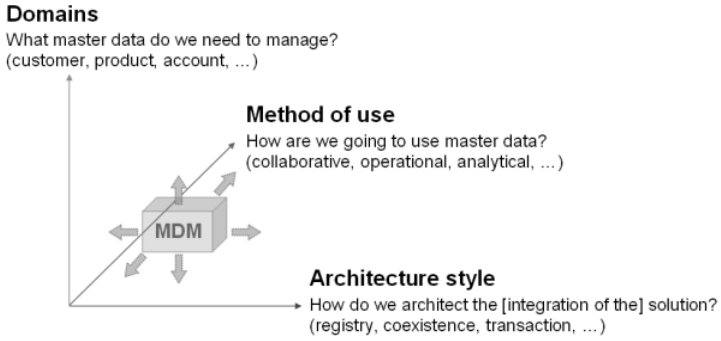

Figure 2.2: Three dimensions of master data management

Master Data Domains

Master data management has emerged over the past few years from the recognition that the existing markets of customer data integration (CDI) and product information management (PIM) had key similarities as well as differences. CDI focuses on managing people and organizations — which we will collectively call *parties*. PIM systems manage the definition and lifecycle of a finished good or service. CDI and PIM both represent a common pattern — that of aggregating data from existing systems, cleaning and augmenting that data, and then distributing the data to downstream systems.

As CDI and PIM products have matured, it was also observed that while CDI systems focused on the customer, it was often convenient for such systems to include references to the products or accounts held by a customer. Similarly, PIM systems often need to store or reference the suppliers of products or materials. Supporting and using these cross-domain relationships has become a key aspect of MDM systems.

In general, we can categorize master data according to the kinds of questions they address. Three of the most common questions — who, what, and how — are addressed by the *party*, *product*, and *account* domains of master data. *Location* information is tied to a product, a party, or an account; it does not have an independent existence in this context.

Figure 2.3 depicts how the three primary domains of party, product, and account overlap. These areas of overlap are particularly interesting because they indicate fundamental relationships between the domains. Explicitly capturing these relationships within the same environment lets us address business questions that may otherwise be difficult to resolve. Indeed, perhaps the key benefit of supporting multiple

domains of master data within the same system is that it clarifies these cross-domain relationships.

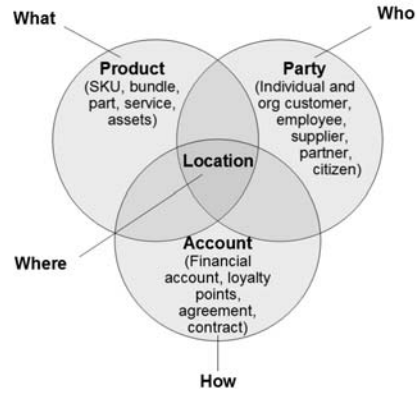

Figure 2.3: Domains of master data

Methods of Use

As we look at the roles master data plays within an organization, we find three key methods or patterns of use: collaborative, operational, and analytical, illustrated in Figure 2.4.

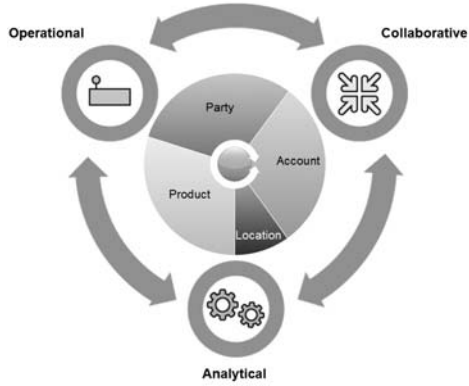

Figure 2.4: MDM methods of use

The simplest way to think about these methods of use is to consider who will be the primary consumers of the master data. Under the *collaborative* pattern, the MDM system coordinates a group of users and systems to reach agreement about a set of master data. Under the *operational* pattern, the MDM system participates in the operational transactions and business processes of the enterprise, interacting with other application systems and people. Last, under the *analytical* pattern, the MDM system acts as a source of authoritative information for downstream analytical systems and sometimes is a source of insight itself. A particular element of master data, such as a product or an account, may be initially authored using a collaborative style, used operationally by other applications, and then also used by analytical systems for decision support.

It is important to note that the style of use is completely independent of the domain of information managed. Although PIM systems often are associated with a collaborative style of use and CDI systems with an operational style, this alignment is not necessary or exclusive. Organizations are increasingly seeking an operational

use of product information as well as a range of use cases for collaborative authoring of customer information.

Collaborative MDM is about achieving agreement on a complex topic among a group of people. The process of getting to agreement is often encapsulated in a workflow that may incorporate both automated and manual tasks, both of which are supported by collaborative capabilities. Information about the master data being processed is passed from task to task within the workflow and is governed throughout its lifecycle. A common set of services is typically also provided to enforce security and privacy and to support administration, validation, and import/export of master data. These services are needed across all kinds of MDM systems.

In the operational style of MDM, the MDM server acts as an online transaction processing (OLTP) system that responds to requests from multiple applications and users. Operational MDM focuses on providing stateless services in a high-performance environment. Operational MDM services are often designed to fit within a Service Oriented Architecture as well as in traditional environments. A wide range of capabilities is required for the operational use style. There can be hundreds of services that provide access and management of MDM data. Specific sets of services for each kind of managed MDM object provide for creation, reading, updating, and deleting the MDM objects. Services are also provided to relate, group, and organize MDM objects. As with the collaborative style of MDM, services are also needed for cleansing and validation of the data, detection and processing of duplicates, and managing the security and privacy of the information.

Analytical MDM is about the intersection between business intelligence and MDM. BI is a broad field that includes business reporting, data warehouses, data marts, data mining, scoring, and many other fields. To be useful, all forms of BI require meaningful, trusted data. Increasingly, analytical systems are also transitioning from purely decision support to more operational involvement. As BI systems have begun to take on this broader role, the relationship between MDM systems and analytical systems has also begun to change. One of the common drivers for clean and consistent master data is the need to improve the quality of decision-making. Using an MDM system to feed downstream BI systems is an important and common pattern. The data that drives a BI system must be of a high quality if the results of the analytical processing are to be trusted. For this reason, MDM systems are often a key source of information to data warehouses, data marts, online analytical processing (OLAP) cubes, and other BI structures. Insight may also be derived from data maintained by the MDM system itself.

An MDM system contains all the information needed to report on key performance indicators, such as the number of new customers per week, the number of new accounts per day, or the average time to introduce a new product. Reporting and dashboarding tools can operate directly over the master data to provide these kinds of domain-specific insights. Analytics may also be executed as an MDM transaction is taking place, using architected integration points that let external functions be invoked as part of an MDM service. A good example is the use of scoring functions to

predict the likelihood of a customer canceling accounts at an institution. This kind of analytics, called *in-line analytics* or *operational analytics*, is an important new way in which MDM systems can work together with BI systems to provide added value to an enterprise.

In practice, MDM usage will often cross the boundaries between collaborative, operational, and analytical use. For example, collaborative MDM processes can be very useful in managing the augmentation of complex operational structures, such as organizational hierarchies. On the other hand, valuable analytical information can be gathered around the nature of the collaborative processes. An MDM implementation may start with the use style that is most important to achieving the business need and then later extend the environment to incorporate additional styles to meet further requirements.

MDM Implementation Styles

Organizations implement MDM systems to improve the quality of master data and to provide consistent, managed use of this information in what is often a complex and somewhat tangled environment. Different combinations of implementation and use requirements have led to the evolution of several MDM implementation styles. Hybrid implementations that combine multiple implementation styles are common. Because some styles are simpler than others, organizations may start with a simpler implementation style that addresses the most urgent business needs and then subsequently address additional needs by extending the implementation to enable additional styles.

In this section, we introduce four common implementation styles: consolidation, registry, coexistence, and transactional hub. As the styles progress from the consolidation implementation style to the transactional hub style, they provide increasing functionality and also tend to require more sophisticated deployments.

The *consolidation* implementation style brings together master data from a variety of existing systems, both databases and application systems, into a single managed MDM hub. Along the way, the data is transformed, cleansed, matched, and integrated to provide a complete "golden record" for one or more master data domains. This golden record serves as a trusted source to downstream systems for reporting and analytics or as a system of reference to other operational applications. Changes to the data primarily come in from the systems that feed it; this is a read-only system. Figure 2.5 illustrates the basic consolidation style, with reads and writes going directly against the existing systems and the MDM system (in the middle) receiving updates from these existing systems. The integrated and cleansed information is then distributed to downstream systems (e.g., data warehouses) that use, but don't update, the master data.

A strong similarity exists between the consolidation implementation style and an operational data store (ODS), which also serves as an aggregation point and staging area for analytical systems such as data warehouses. The distinction lies in the set of

platform capabilities that an MDM system offers, which go beyond the storage and management of data that an ODS provides. An operational data store is a database that is used in a particular way for a particular purpose, while an MDM system provides access, governance, and stewardship services to retrieve and manage the master data and to support data stewards as they investigate and resolve potential data quality issues.

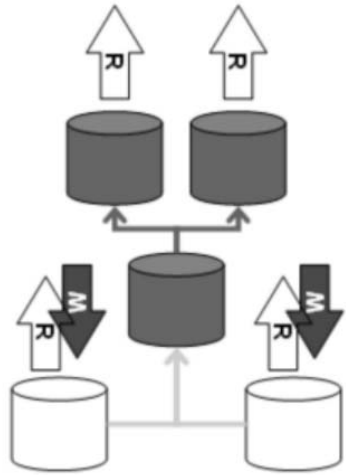

Figure 2.5: Consolidation MDM implementation style

Implementing the consolidation style is a natural early phase in the multiphase roll-out of an MDM system. A consolidation style MDM system serves as a valuable resource for analytical applications and at the same time provides a foundation for the coexistence and transactional hub implementation styles.

The drawbacks of the consolidation style mirror its advantages. Because it is fed by upstream systems, it does not always contain the most current information. If batch imports are performed only once a day, the currency requirements for a decision support system will likely be met — but those for a downstream operational system may not be. Because the consolidation style represents a read-only system, all the information about a master data object must already be present in the systems that feed the MDM system. Thus, if additional information needs to be collected to address new business needs, one or more of the existing source applications must be changed, as well as the MDM system. The coexistence and transactional hub implementation styles address this lack of flexibility.

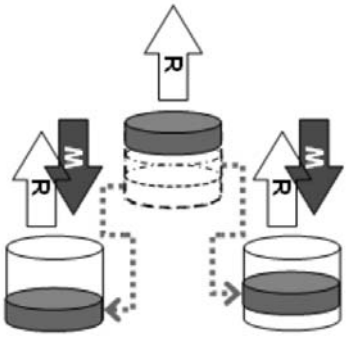

Figure 2.6: Registry MDM implementation style

The *registry* implementation style can be useful for providing a read-only source of master data as a reference to downstream systems with a minimum of data redundancy. In Figure 2.6, the two outside systems are existing sources of master data. The MDM system in the middle holds the minimum amount of information required to uniquely identify a master data record; it also provides cross-references to detailed information managed within other systems and databases. The registry is able to clean and match just this identifying information and assumes that the source

systems are able to adequately manage the quality of their own data. A registry style of MDM implementation serves as a read-only system of reference to other applications.

Queries against the registry style MDM system dynamically assemble the required information in two steps. First, the identifying information is looked up within the MDM system. Then, using that identity and the cross-reference information, relevant pieces of information are retrieved from other source systems — a process called federation. Federation can be done at the database layer or by dynamically invoking services to retrieve the needed data in each of the source systems. The registry implementation style can also be useful in complex organizational environments where one group may not be able to provide all its data to another. Because responsibility for most of the data remains within the source systems, the registry style can be relatively quick to implement.

However, a corresponding set of issues exists with this implementation style. One fundamental issue is that a registry implementation is not useful in remediating quality issues that go beyond basic identity. A registry implementation can manage the quality of only the data it holds — so while it can match and cleanse the core identifying data, it cannot, in itself, provide a completely standardized and cleansed view of the master data. Because the complexities of updating federated information lead most registry style implementations to be read-only, the cleansed identifying information is not typically sent back to the source systems. If the data in the source systems is clean, the composite view served by the MDM system will also be clean. Thus, a registry implementation can act as an authoritative source of master data for the key identifying information it maintains.

A registry implementation style is also more sensitive to the availability and performance of the existing systems. If one of the source systems slows down or fails, the MDM system will be directly affected. Similarly, the registry style requires strong governance practices between the MDM system and the source systems because a unilateral change in a source system could immediately cause problems for users of the MDM system.

The *coexistence* style of MDM implementation involves master data that may be authored and stored in numerous locations and that includes a physically instantiated golden record in the MDM system that is synchronized with source systems. The golden record is constructed in the same manner as in the consolidation style, typically through batch imports, and can be both queried and updated within the MDM system. Updates to the master data can be fed

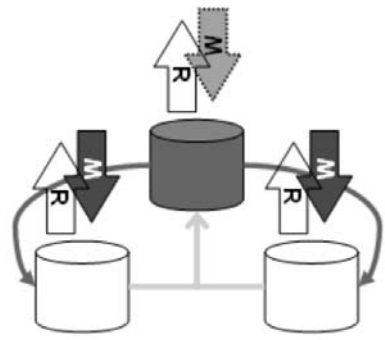

Figure 2.7: Coexistence MDM implementation style

back to source systems as well as published to downstream systems. In a coexistence style, the MDM system can interact with other applications or users, as shown in Figure 2.7.

An MDM system implemented in the coexistence style is not a system of record, because it is not the single place where master data is authored and updated. It is a key participant in a loosely distributed environment that can serve as an authoritative source of master data to other applications and systems. Because the master data is physically instantiated within the system, the quality of the data can be managed as the data is imported into the system. If the MDM system does a bidirectional synchronization with source systems, care must be taken to avoid update cycles where changes from one system conflict with changes from another — these cycles can occur through a combination of automated and manual conflict detection and resolution.

The advantage of the coexistence style is that it can provide a full set of MDM capabilities without causing significant change in the existing environment. The disadvantage is that because it is not the only place where master data may be authored or changed, it is not always up-to-date. As with the consolidation style, the coexistence style is an excellent system of reference but is not a system of record.

A *transactional hub* implementation style (Figure 2.8) provides a centralized, complete set of master data for one or more domains. It is a system of record, serving as the single version of truth for the master data it manages.

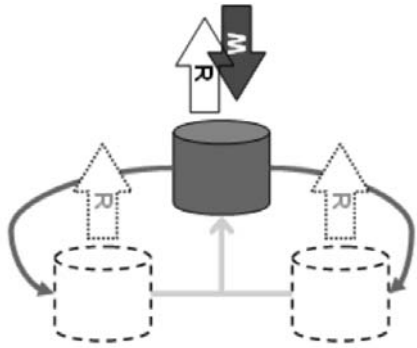

Figure 2.8: Transactional hub MDM implementation style

A transactional hub is part of the operational fabric of an IT environment, receiving and responding to requests in a timely manner. This style often evolves from the consolidation and coexistence implementations. The fundamental difference is the change from a system of reference to a system of record. In the system of record, updates to master data happen directly to this system using the services provided by the hub. As update transactions take place, the master data is cleansed, matched, and augmented to maintain the quality of the master data. After updates are accepted, the system distributes the changes to interested applications and users. Changes can be

distributed as they happen via messaging or can be aggregated and distributed as a batch.

Sometimes, an MDM system requires data extensions to accommodate information that is not already stored in the source systems. Rather than augment all the source systems, the MDM system is extended to support this new information and becomes the only place where such information is managed.

Governance and security are key aspects of all MDM implementation styles. Access to the master data must be tightly controlled and audited. Auditing can be used to track both queries and changes to the data. Visibility of the information can be controlled to the attribute value level to ensure that the right people and applications are restricted to seeing the right information in the right context. Because a transactional hub implementation is a system of record, security and governance play an especially critical role in maintaining the integrity of the master data.

The benefits of a transactional hub implementation are significant. As the system of record, it is the repository of current, clean, authoritative master data, providing both access and governance. Any of the methods of use (collaborative, operational, and analytical) can be implemented to meet the MDM needs of an organization. The primary difficulty in a transactional hub implementation is achieving the transition from system of reference to system of record. As a system of record, all updates should be funneled to the MDM system; this means that existing applications, business processes, and perhaps organizational structures may need to be modified to use the MDM system. Although this change is potentially costly, the overall organization generally benefits as more comprehensive data governance policies are established to manage the master data.

The primary disadvantages of the transactional hub style are cost and complexity. The implementation of a transactional hub often means that existing systems and business processes must be altered when the transactional hub becomes the single point of update within the environment. To minimize disruption, you can perform the transition to a transactional hub incrementally. The significant benefits of a transactional hub implementation cause it to be the ultimate goal of many MDM projects.

The different implementation styles introduced in this section are complementary and additive. You can implement different MDM domains with different styles within the same MDM system. As we have mentioned, it is common for an MDM deployment to start with one style, such as the consolidation style, achieve success with that implementation by publishing authoritative master data to downstream systems, and then extend the system with a coexistence style. With the completion of the coexistence phase, the MDM system could then be used to support the master data needs of new applications while continuing to publish snapshots of master data to downstream systems. Over time, the existing systems could be altered to leverage the MDM system, which would become a system of record. In an MDM system supporting multiple domains of master data, such as customer, product, and supplier, we

may find that the MDM system may appear as a consolidation style for one domain, a registry style for a second domain, and a transactional hub for a third domain.

Enterprise Information Models

Next, we look at the role that information modeling plays in addressing the business challenges raised by the need for trusted information. Specifically, we are interested in the delivery of trusted information, but the information of interest typically is scattered across a range of different systems, with very different structures, data quality characteristics, and business semantics. Given this environment of disparate legacy systems, how do we integrate this information? We know it is not enough simply to "copy" data from multiple sources, as discussed above. We need a way to understand the information we have and a way to understand how to structure it for the delivery of trusted information.

To address these needs, we introduce the concept of *enterprise information models* that address different levels of information modeling with a goal of standardization across multiple competing perspectives, such as line of business or channel. Information models will, of course, exist at a number of different stages in the analysis–design lifecycle. Figure 2.9 identifies the model types involved.

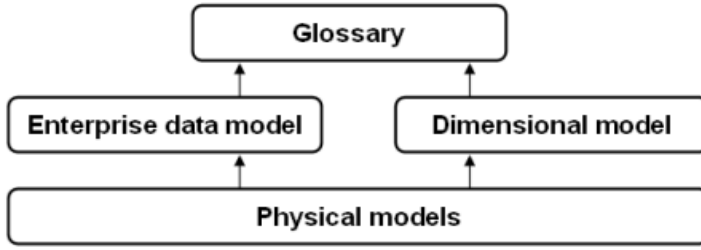

Figure 2.9: Enterprise information models

The first model type is a standard lexicon of business terminology — often called a *glossary*. The goal of this model is to drive standardization of the terms used in expressing business requirements across the expression of information concepts (such as "product type" or "customer tax status"), business functions ("risk management" or "arrangement negotiation"), reporting requirements ("total cost of outward customer communications"), or even activities and verbs ("activate"). This model type is key to standardization of business language with enterprise information models and provides the terminology upon which subsequent model types build.

A second model type is a normalized expression of the primary information entities of interest to the business. This is often called *an enterprise data model (EDM)*. An EDM is typically a logical data model — one that does not capture the specifics of any particular data store but rather abstracts from the physical implementations of

the data platform to provide a logical view of the constructs involved and how they relate across the enterprise. The EDM is often mapped to the business glossary so that concepts in the EDM can be related back to the standard business language of the enterprise. The EDM is key to the rationalization of multiple information sources, as well as the specification of new information sources, such as a data warehouse. Physical data models (PDMs) which describe the actual format of a data source, can be usually be generated from logical data models by entity relationship (ER) modeling environments.

The final model type of interest in this section is a *dimensional model*, which details the measures and dimensions of trusted information that are exposed to consumers. Dimensional models are typically related to the EDM in that information from the EDM is extracted and restructured to form a *data mart* — a data store that exposes enterprise information in a way that is efficient for business reporting. Dimensional models also relate to the glossary model.

There are, of course, many more types of enterprise information models, including service models, message models, and even business process models. However, the goal here is to highlight those model types that are of most relevance to the remaining topics in this chapter.

IBM's *Industry Models* are a collection of interrelated enterprise information models addressing different aspects of the analysis and design of trusted information for key industries, such as banking (Information FrameWork, or IFW), insurance (Insurance Application Architecture, or IAA), healthcare (Health Plan Data Model, or HPDM), telecommunications (Telecommunications Data Warehouse, or TDW), and retail (Retail Data Warehouse, or RDW). In each model set, we encounter these same model types: a glossary, an enterprise data model (focused on the specification of a data warehouse), and a set of conformed dimensional models.

The goal of IBM's Industry Models is to deliver that intellectual property within an industry that ports well between organizations, across geographies, and across regulatory domains. As such, the Industry Models do not provide an "out of the box" solution to the delivery of trusted information. Instead, they are intended to capture industry-specific constructs to accelerate the analysis and design efforts of an organization seeking to deliver that trusted information.

Standardizing Business Language

We have established that the need for trusted information introduces a range of business challenges. One such challenge is the need to drive a consistent understanding of information across the enterprise, eliminating the natural variance that is introduced through human communication; this is the role of a glossary. A glossary is typically deployed across the enterprise through a set of data stewards. These stewards are responsible for ongoing maintenance of the terms in the glossary and for ensuring that the terms and their definitions are aligned with the business.

Earlier, under the topic of business challenges, we identified alignment of business and IT as a key barrier to the delivery of trusted information. A glossary represents the first step in bridging this gap, by enabling business and IT stakeholders to speak the same language. By providing detailed business definitions, aliases, and related terms, we can reduce the likelihood that the same term is interpreted in multiple ways, leading to misalignment of the semantics of information. This enables the common business vocabulary that we identified the need for earlier in this chapter. Structuring this vocabulary as an enterprise business glossary with appropriate data stewardship will help with the penetration of the vocabulary across the organization, in both the business and IT communities.

The glossary also plays a role in understanding the available information within the enterprise and taking action to address data quality. In discussing information integration, we described the use of data profiling capabilities to analyze existing data sources to ascertain data quality. Without some reference point of business language, this task is, at best, challenging. How do we go about assessing the relative data quality of customer data within legacy systems across the enterprise if we do not have an agreed understanding of what "customer" means? Delivery of trusted information hinges on a good understanding of the meaning of that data. An enterprise business glossary is key to that understanding.

InfoSphere Information Server delivers this capability through Business Glossary, which provides secure access and governance of business terms across the enterprise. These terms are then leveraged across the InfoSphere Information Server to classify IT artifacts according to their business meaning. Source systems, ETL specifications, quality transformations, OLAP structures, and even report fields are traced to the business terms they expose. Irrespective of the differences in terminology used within the physical assets, we can identify all cases where a given business concept is used and can trace a given concept through a business intelligence solution.

Formalizing the Requirements Process

Enterprise information models play a role in formalizing the gathering of business requirements. When no model exists, the result is a void in which multiple competing mechanisms for the expression of business requirements (including no formal mechanism at all) will flourish. This void rapidly leads to a misinterpretation of the goals of business stakeholders and to ambiguity as to the intent of information constructs. This situation is what leads to the reinterpretation of closely related data structures over time, resulting in poor data quality and duplication of related information.

Through the use of an enterprise information model, the requirements-gathering process can be formalized. We can now express business requirements through the formalized language of a glossary, eliminating much of the ambiguity of the initial requirement specification. This requirement specification can then be mapped to the

EDM constructs it impacts, further formalizing the requirement against a concrete set of structures.

Enterprise models also encourage a separation between analysis and design — a distinction critical to a successful requirements-gathering process. Such a separation enables us to decouple discussions about what information we wish to have from discussions about what is actually implemented today. Without this distinction, requirements discussions inevitably become tainted by the data structures in place today, which severely limits the ability of the organization to capture its true information requirements. Using an enterprise model encourages the use of separate logical data models and physical data models to capture analysis and design constructs, respectively.

InfoSphere Data Studio enables such a requirements process through the capabilities of its data modeling environment. With this tool, the glossary, enterprise data model, and dimensional models can be formally managed and mapped to requirements documents (e.g., within Rational Requisite Pro). In the case of IBM's Industry Models, a formalized methodology is added to further assist with the requirements-gathering process, guiding practitioners with leading practices that have been proven in the field.

Integrating Information Sources

The information required by data warehousing and business intelligence initiatives is rarely available in the required form. It typically is spread out over the enterprise, with different structures, formats, and often conflicting business meaning. A key challenge to the delivery of trusted information is to enable an understanding of these data sources so that existing information can be used consistently and accurately.

Within the context of data transformation and data cleansing, we need a way to understand the distribution of business concepts across these sources and a way to use that understanding to accelerate the development of a data integration solution.

As we have seen, a business glossary is central to driving this common business understanding. It provides us with a formal set of business terms we can use to drive an understanding of existing sources using a common, agreed-upon business language. Given this understanding, we can start to understand source-to-target mappings. The glossary definition is part of an enterprise information model, and IBM provides these definitions as part of the Industry Model offerings. The reason for obtaining this understanding indirectly, by mapping everything to a common business language, is largely a factor of the potential scale of mapping involved. Although on a small scale it may seem feasible to map source to target directly, this approach becomes less appropriate as the scale of the data integration grows. Without any single defined business language to guide the analyst, maintaining a clear understanding of the business concepts involved becomes very difficult. Instead, the mapping exercise will become embroiled in the details of each physical data implementation.

A glossary model and the classifications of physical assets can be published to or constructed within InfoSphere Information Server. Given this metadata, InfoSphere FastTrack lets the analyst rapidly define source-to-target mappings based on these glossary classifications. This results in an analysis specification of the data integration (ETL) requirements, including transformation instructions, which can be used to generate initial DataStage job definitions. This specification provides critical guidance to the ETL developer, letting that developer complete the DataStage job specification.

Enterprise models also guide the development of the target schema involved in the data integration project. Based on the business terms in the glossary, we can infer the parts of the EDM that are of interest in a particular project. This lets us derive a logical data model specific to the proposed target of the information integration (e.g., a data warehouse). This logical data model can in turn be used to derive a physical data model and the data definition language (DDL) necessary to construct the target datastore.

Defining a BI Solution

Given these distinct applications of an enterprise information model, it becomes clear that such a model plays a key role in the delivery of trusted information such as a BI solution. This section summarizes the steps involved in the development of such a solution, with an emphasis on the role of the enterprise information model.

First, let us clarify the landscape under discussion. As Figure 2.10 illustrates, key business information is contained within a known set of source systems, in varying metadata structures, with varying levels of data quality. We want to extract and transform the data from these sources into a new target data warehouse. From this warehouse, we then extract a set of data marts that can be exposed, through OLAP, to the information consumer in the form of a report.

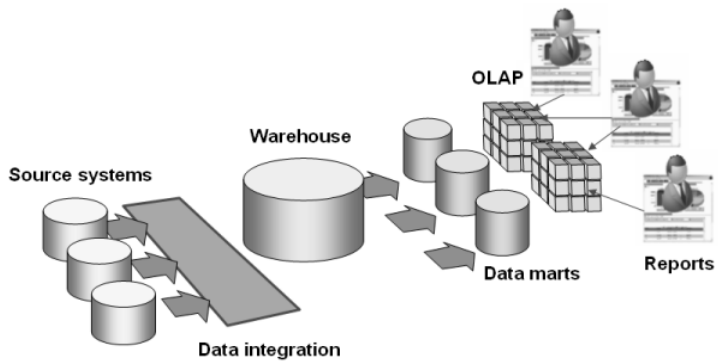

Figure 2.10: A typical BI landscape

Given this landscape, how do we ensure that the information we deliver to the report consumer is of sufficient quality and can be trusted, and what does this mean for the enterprise information model?

First, we must understand the information sources that exist today. As we have discussed, this task involves a number of distinct dimensions of analysis. We must understand what business information is represented within each source, an understanding we achieve by classifying the physical data elements using the terms of a glossary. Second, we must understand the quality of the information within each source. This is the role of data profiling and analysis, activities in turn informed by an understanding of the data concepts through the glossary. Last, we must understand the relationship and overlap between these sources, knowledge we obtain by assessing where the same glossary terms classify multiple different physical data assets.

Given this understanding of source systems, we look to define a new integrated target datastore — in this case, a data warehouse. This definition is driven largely by the enterprise data model and typically is accomplished iteratively based on the systems being integrated. Based on the glossary terms that classify the source systems and on the relationship between these terms and our EDM, we can identify the subset of the EDM that is of interest in our current project. This EDM subset is extracted as a warehouse logical data model. Based on this warehouse logical model, we can define a physical model and generate the DDL necessary to create the warehouse on our target database platform.

We now have an understanding of our data sources and an understanding of our proposed target, expressed in terms of a consistent glossary. Based on this knowledge, we can model the ETL specifications necessary to load the warehouse from the source systems. This task can be greatly accelerated through the use of a logical mapping environment such as InfoSphere FastTrack, which can then generate initial InfoSphere DataStage jobs for completion by the ETL developer. The ETL developer also takes the outcome of the data profiling and quality analysis into account in defining the data transformation and cleansing.

Trusted information is now available within our warehouse system of record, but we have not yet enabled its delivery. The data warehouse (system of record) typically is not structured for efficient reporting schemas; its role is to aggregate and hold the required data in a way that minimizes duplication and redundancy (i.e., normalizes) in preparation for extraction to a set of data marts. These data marts are typically structured to enable rapid extraction of measure and dimension sets by an OLAP environment and subsequent presentation through a business report. The specification of the enterprise information model includes a dimensional model that guides the specification of these data marts. Ideally, the specification of the data marts is developed from an enterprise perspective, as opposed to data marts being defined “as needed.” This practice eliminates duplication and redundancy between data marts and lets the marts grow in a controlled way over time, rather than organically.

The data extraction from the warehouse system of record to the data marts can be performed in a similar way to the loads from the sources to the system of record. Recall that both the EDM and the dimensional model are classified by a common business glossary. This mapping is maintained when the physical models are generated. This means we can again rapidly understand the source–target mappings and transformations required to populate our data marts.

We can now import these data mart specifications into a BI environment, such as Cognos, where OLAP structures can be linked to report user interfaces to deliver information to business users.

The shared metadata platform of InfoSphere Information Server now enables some interesting side effects of building such a solution. Because the physical metadata for each part of this solution (sources, ETL, warehouse system of record, data marts, and even OLAP and BI structures) is available within the metadata repository, we can perform full data lineage through this chain, answering key questions such as "What business reports expose this data?" or "Where did the data on this report come from?" In addition, because each element in this chain is mapped to a common business glossary, we can ensure a common semantic understanding of each link in this lineage and indeed perform impact analysis, answering such questions as "If I change this ETL job, which reports will I break?"

Model Management

Of course, the maintenance of enterprise information models is not trivial. By their nature, these models are very broad in their business scope and express concepts that cross lines of business and other organizational boundaries. It is not always in the interest of each individual shareholder to share information or metadata in this way. For example, it is not always in the best interest of one line of business to share its customer relationships with another. A significant barrier exists to the delivery of trusted information in the successful management of the enterprise information model itself.

This challenge typically requires a dedicated organization unit to manage the model and its relationships to information sources. This does not imply that this unit is solely responsible for driving the enterprise-wide understanding of all business information; this understanding is almost always best evolved through distinct projects with a defined goal. However, this organization unit must exist. Its role is to act as custodian of the enterprise information model, to ensure that appropriate standards are maintained, and to facilitate the model's use by integration projects within the enterprise.

3

Managing Information: Dynamic Warehousing

Since the introduction of the very first relational databases, IBM has been providing industry-leading platforms for delivering analytics to business users. Over time, these systems have been known as everything from executive information systems to enterprise data warehouses. One thing, however, has remained the same: the ability to deliver enterprise-class solutions to both small and large businesses alike from IBM.

During the early days of reporting solutions, much reporting took place from within applications, which often sat on varying technology platforms. As time and technology progressed, organizations began to consolidate data from many different platforms into centralized, purpose-built reporting solutions, or data marts. As these solutions gained popularity and number throughout organizations, the issues of consistency in reporting and spiraling cost of ownership brought the advent of centralized data reporting environments. Those environments have evolved into what we now know as data warehouses.

As the thirst for business intelligence increases, the need for timely information quickens. Businesses no longer want to wait until information filters into the warehouse overnight. They'd rather see data arrive in the warehouse in near real time so they can make informed decisions on the spot to perform in the dynamic business environment of today. This requirement includes the ability to embed warehouse-driven analytical services into day-to-day business operations.

Traditional data warehousing approaches play a critical role in enabling companies to review historical data in order to evaluate business performance for strategic and tactical planning purposes. Moving forward, the fast-changing competitive landscape places additional requirements on warehousing solutions to enable access to multiple types of data and provide real-time views of business operations to a broader audience.

Several industry trends are driving the evolution of the modern data warehouse:

- Leveraging all data assets through consolidated and cross-functional systems (single view of the truth)
- Extending the use of data analysis into core operations and business processes (operational business intelligence)
- Moving toward real-time or "right-time" data to support business decision-making (requiring reduced data latency)
- Moving to "time-to-value" and simplicity through appliance-like deliveries

What Is Dynamic Warehousing?

To achieve the vision of a single view of information within an organization, businesses increasingly have to invest in information integration and warehousing strategies to address timeliness of data and data quality needs. They also need to integrate their business analytics into operational applications so that fraud detection, customer churn, and business exposures can be identified before they occur (as opposed to retrospectively), thus improving profits and business efficiency. Enter the age of *dynamic data warehousing*, where analytics is no longer about history but about having up-to-date information that enables businesses to make timely decisions to strengthen productivity, reduce risks, and increase opportunities to improve revenue.

As we move into this new age of dynamic data warehousing, with the demands of real-time data access and user consolidation, a robust integration and warehouse architecture is required. Also necessary is a common framework that defines common terms, models, and usage for how information within a business relates across the warehouse and source.

Among the challenges that need to be addressed by dynamic warehousing for real-time or right-time workloads are

- Managing mixed workloads
- Supporting 24x7 availability (i.e., no batch window)
- Minimizing data latency to support applications requiring streaming analytics and event processing (e.g., stock trading, radio frequency identification, compliance, network analysis)

By increasing speed or response time, we can improve the organization's decision-making ability. Data latency can take many forms — the longer it takes to extract, load, or transform the data before it becomes usable in the data warehouse, the less responsive and effective the organization's decision-making. Dynamic warehousing applications often receive data continuously and need to manage the varying demands for resources, whether they be large or small queries or even quick update requests.

Increasingly mixed workload environments and the constantly changing needs of different business constituents require more dynamic warehousing capabilities than ever before. No longer do data warehouses serve just traditional query and reporting purposes. Today's dynamic warehouses must

- Optimize performance for increasing demands to manage more data
- Simplify operations and the effort required to integrate with other tools
- Leverage all types of information, including unstructured information (e.g., text, XML)
- Serve increasing numbers and types of applications and users, with varying service level demands

Why Is Dynamic Warehousing So Important Moving Forward?

Dynamic warehousing requires a new approach to information management: one that enables real-time analysis of all information types in support of process optimization and risk reduction; one that helps generate more knowledge about customers to identify problems and improve service; and one that helps address information requirements driven by governmental regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), Sarbanes-Oxley, and Basel II. The ultimate goal is to free information from the silos in which it resides so it can be delivered as a service, in the right context and format, to any business process or user on demand.

IBM believes dynamic warehousing is the next logical step in the evolution of data warehousing. A dynamic warehousing approach makes it easier for IT organizations to support business requirements for actionable information — not just data, but all types of information, with intelligence behind it to help people take action and make decisions. Unlike historical data warehouses and business intelligence approaches, dynamic warehousing delivers immediate, integrated information; it empowers users by unlocking the value of information and analytic capabilities; and it can service multiple applications and business lines for both strategic planning and operational purposes with a flexible architecture.

Dynamic warehousing is not a product, a tool, or a simple one-off solution. It is an approach that enables you to deliver more dynamic business insights by integrating, transforming, harvesting, and analyzing insights from structured and unstructured information. The result is a framework for delivering right-time, contextual information for both strategic planning and operational purposes.

Enabling dynamic warehousing requires a set of services that extends beyond traditional data warehousing and reporting to address the demands for more dynamic business insight and support the increasing number of business processes and applications requiring analytic capabilities. Table 3.1 lists some examples of how dynamic warehousing can deliver enhanced insights and results.

Table 3.1: Dynamic warehousing examples	
Traditional warehousing	**Dynamic warehousing**
Insurance fraud analysis and reporting	Identifying potentially fraudulent claims before approval and payment *— Transforms healthcare*
Reporting on customer issues	Identifying possible related issues, churn risk, and cross-sell opportunities while engaged with the customer *— Transforms customer service*
Historical sales analysis and reporting	Discovering relevant customer information to identify cross-sell opportunities and improve negotiating position at the point of sale *— Transforms sales effectiveness*
Crime statistics and reporting	Identifying related incidents and potential suspects before arriving at the crime scene *— Transforms crime fighting*

The Enterprise Data Warehouse Appliance

The *IBM InfoSphere Balanced Warehouse* (Figure 3.1) is more than a fixed-function appliance, providing unique characteristics. While it is prescriptive and appliance-like in nature, it can also easily be incorporated into an existing IT environment through the use of widely used technologies and the ability to integrate with standard business operational infrastructure for operations such as backups, monitoring, and system management.

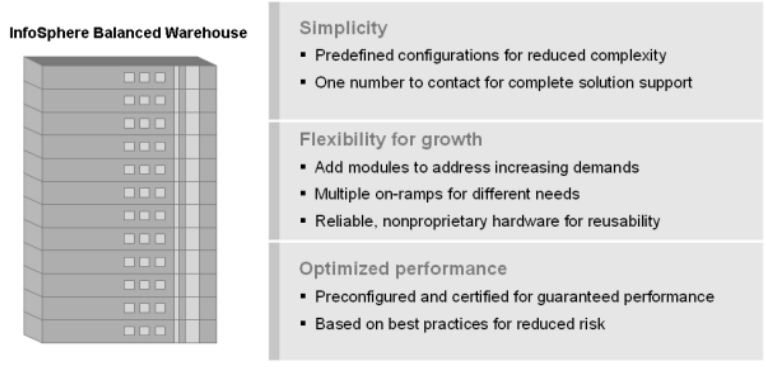

Figure 3.1: IBM InfoSphere Balanced Warehouse

For any business, the difference between receiving a ready-to-go system, such as the InfoSphere Balanced Warehouse, versus one that needs to be fully or partially built on arrival can be dramatic in terms of time-to-value and project risk. A pre-built InfoSphere Balanced Warehouse platform offers the following key advantages over a build-your-own environment:

- Reduced risk
  - » Pre-built to your requirements from a proven and tested solution
- Reduced time-to-value
  - » Rapid deployment
  - » Predictable performance
  - » Easy integration into the IT environment
- Reduced warehouse cost of ownership
  - » Reduced management costs due to automated management
  - » Simplified problem resolution and system upgrades due to single product support
- Affordable growth aligned to business requirements
  - » Incremental approach to growth at a predictable cost

As the demand for resources increases, companies will require more balanced and optimized performance from warehouses (balanced storage, hardware and software performance) to keep costs in check and meet varying service level requirements. The InfoSphere Balanced Warehouse provides a modular architecture that enables granular growth or can grow incrementally as the business requires.

The modular architecture was specifically designed to represent a simpler, more transparent architecture, uncommon in today's warehousing appliance market. Pre-balanced building blocks give you the freedom and flexibility to decide your warehousing growth path. A preconfigured, pretested allocation of software, storage, and hardware supports a specified combination of function and scale.

If the dynamic warehousing application requires more users or data to support, it is a straightforward process to determine the additional modules needed to support the added workload. Last, this architecture is open, and modules can be defined to support third-party applications.

Leveraging, Monitoring, and Managing the Data Warehouse

The choice of warehousing platform can have a major impact on the success, cost, and ongoing viability of your data warehouse. Today, many different warehouse platforms make claims about their capability and low cost of ownership, yet very few can actually provide the level of flexibility, functionality, scalability, and reliability required from a business-critical data warehouse.

The hub of dynamic warehousing is still the data warehouse. Moving forward, however, data warehouses will need to support increasingly mixed application

requirements, including mission-critical operational activities in addition to traditional back-office reporting. Warehouses will also need to provide expanded functionality and to work seamlessly with the other services required to enable dynamic warehousing. A data warehouse solution should be able to consolidate data marts and siloed solutions across the enterprise into a unified warehouse where needed, while still enabling distributed data marts to address business-specific requirements.

IBM InfoSphere Warehouse is software integrated with the InfoSphere Balanced Warehouse solution that both small and large businesses can use to gain the insight and information agility they need more quickly, simply, and efficiently than ever before. With this complete, multipurpose environment, organizations can access, analyze, and act on real-time operational and historical data to consistently generate new opportunities, contain costs, and satisfy customers with minimal business risk. Its simplicity, flexibility, and reliability provide an infrastructure that enables it to meet business needs, whether in terms of functionality or to address a growing volume of information or users.

It begins with a massively parallel processing database that can address the performance requirements of your application. The massively parallel architecture is an important foundation for the InfoSphere Warehouse because it enables the warehouse to scale linearly, so that as workload increases, we can accurately predict the necessary resources required to meet the new demands. It is this foundation that also enables the InfoSphere Balanced Warehouse to provide modular growth. A set of integrated tools for data modeling and design, administration, and control of the warehouse and all related services is also included.

Embedding Analytics into the Warehouse

InfoSphere Warehouse extends the value of the warehouse by including analytics and visualization to provide more dynamic business insight. These services are embedded in the warehouse to enable better performance, increased efficiencies, and reduced costs.

Many business users use Microsoft Excel spreadsheets to store mission-critical financial data. Excel gives users the ability to issue multidimensional expressions (MDX), advanced queries that can access data stored in a multidimensional structure. The query speed of these operations is often several times faster than retrieving data from a traditional database. *InfoSphere Warehouse Cubing Services* supports MDX queries from Excel and other business intelligence tools and returns the answer by retrieving the data from its database cache or retrieving it from disk if the data is not found in the cache. BI applications that commonly use these queries include executive information systems, key performance indicators, management reporting, and business performance management. The capabilities provided by Cubing Services offer the opportunity for deeper analysis of business variables to generate new insights from key organizational data (Figure 3.2).

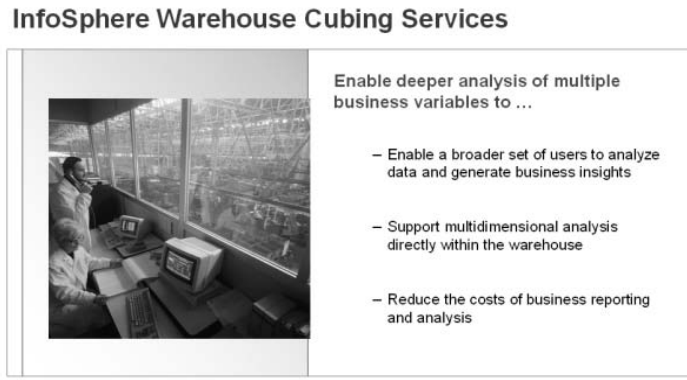

Figure 3.2: Strategic benefits of cubing services

Most data found in organizations today are unstructured — stored in text documents, XML formats, or content management data stores. Leading-edge companies are leveraging such data to improve their decision-making. One way to exploit this kind of data is through *text analytics*.

Unstructured data can't be used "as is" for business intelligence applications. It contains too much "noise," or data that is not relevant. To make text documents usable in a dynamic warehousing environment, we need to extract "structured" information, such as product names, product codes, indicators for problems, or expressions of sentiment, from the unstructured data. This information can then be added to the existing warehouse in additional columns.

InfoSphere Warehouse Text Analytics uses natural language processing technology to extract meaningful, structured information from unstructured text, as well to detect relationships, based on named entities it identifies. With this capability, businesses can derive new attributes and dimensionality from their unstructured information for further OLAP and data mining analysis. An application can target specific information hidden within text, producing a greater return on the application investment and providing a competitive edge by driving further business insight. Examples of the resulting value to the business include better product categorization, early warning on customer attrition, fraud detection, product defect analysis, and better customer profiling.

Predictive analytics, often referred to as *data mining*, differs from traditional BI applications, in which the user must already have some idea what he or she is looking for; in other words, the user has formulated a hypothesis and is trying to prove or disprove it. Predictive analytics lets the user analyze the data without a preconceived hypothesis, using analysis that is data-driven, to discover rules or relations among the data that no one thought to look for. Technology innovators have been able to use data mining to produce high-value results that change how the business competes and invests its assets.

The use of predictive analytics can answer questions such as the following to bring advantages to the business:

- Increase revenue: improve customer satisfaction and retention
  - » What are the characteristics of my high-value customers?
  - » How can I detect opportunities for cross-selling and up-selling?
  - » How can I ensure each store has the right set of products?
  - » Which of my customers are likely to leave?
  - » What is the right insurance quote for this customer?
- Decrease warranty costs: improve the quality of products and services
  - » When one part fails, what other parts may fail soon?
  - » What is the likelihood of a part failure?
- Decrease claim costs: detect potential fraud earlier
  - » Healthcare
  - » Insurance, banking
  - » Government subsidies (e.g., agriculture)

InfoSphere Warehouse supports the popular approaches to performing data mining:

- Discovery mining — Finding patterns in data to guide decisions (association rules, sequential patterns, clustering)
- Predictive mining — Using known results to create models that can predict values (classification, regression)

Embedding analytics within the data warehouse has several advantages, one of which is application performance. Whether the user requires multidimensional or unstructured data or wants to perform data mining, making a copy of the data for the particular tool is not required (Figure 3.3).

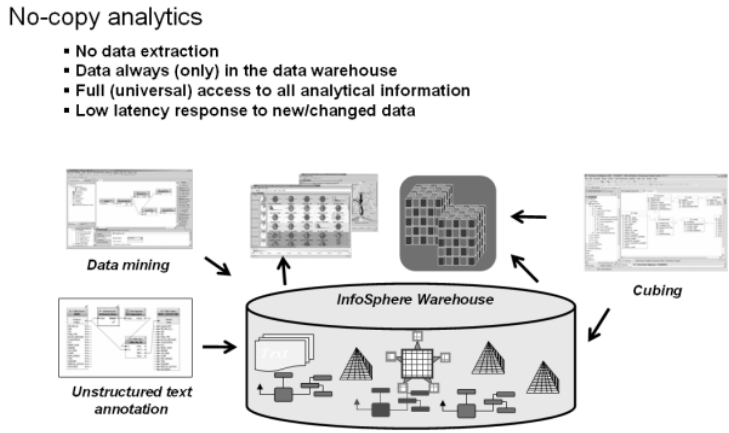

Figure 3.3: Benefits of no-copy analytics

Information Lifecycle Management of the Warehouse

Information lifecycle management is essentially the ability to monitor, analyze, and manage the lifecycle of applications, systems, and processes in the data warehouse environment. The ability to control and optimize this environment includes setting policies to optimize the use of system resources (CPU, memory, disk). By implementing a comprehensive lifecycle management system for the data warehouse, organizations not only can save on costs but also will have a system that can support a high-performance, dynamic warehousing environment.

InfoSphere Warehouse provides several features (Figure 3.4) to support management of the information lifecycle in the warehouse.

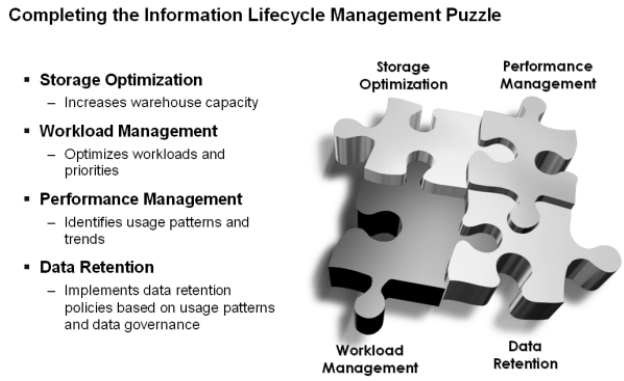

Figure 3.4: IBM InfoSphere Warehouse information lifecycle management features

Storage Optimization

The InfoSphere Warehouse *Storage Optimization Feature* should be considered for any data warehouse implementation. This feature enables the compression of data on disk to decrease disk and storage infrastructure requirements. Along with storage infrastructure savings, the compression capability can provide these benefits:

- Achieve higher performance because compression reduces the number of I/Os required to retrieve the data
- Use memory more effectively because more compressed rows can be held in memory at one time
- Reduce backup images to half their space, and take half the time to finish the backup
- Reduce recovery time due to decreased backup image size
- Reduce cooling, electrical, and space requirements to manage a smaller disk infrastructure
- Reduce network traffic when recovery logs are transmitted to the disaster recovery site

Performance Management

An important consideration when assessing data warehousing appliances or software is the ability to monitor the system, analyze whether any changes occurred, and then take action to optimize or control the environment. The ability to provide improved analysis of the system and enable better problem determination will reduce support costs and maximize performance. The InfoSphere Warehouse *Performance Suite* provides tools to enable organizations to manage, analyze, and control their data warehouse environment:

- InfoSphere Warehouse Performance Optimization Feature — Database- and operating system-level analysis for system-level optimization and control of resource consumption
- InfoSphere Warehouse Performance Monitoring Feature — Query workload analysis that correlates user activity, data usage, and query performance metrics without disrupting or adding overhead to production systems

Data Retention

IBM InfoSphere Warehouse with Optim Data Retention enables businesses to better manage the growth of information by integrating data warehousing and archiving capabilities. This automated, policy-based data lifecycle management solution reduces storage and data retrieval costs and improves performance. Business benefits include the following abilities:

- Improve business insights by querying only the information relevant to the business and archiving (with easy access) the rest
- Effectively manage the size and performance of the data warehouse by archiving information with the confidence that it can be returned to the warehouse if needed
- Ensure accurate, timely response to audit and e-discovery requests, thereby lowering the cost of compliance
- Reduce the cost of managing historical data
- Reduce speed of growth of the data warehouse platform
- Ensure adherence to compliance regulations
- Ensure access to historical data
- Reduce requirements for secondary historical repositories for specific users (e.g., actuaries, analysts)

Workload Management

Companies are consolidating their databases into larger databases both for efficiency and to integrate their enterprise data. These consolidated systems experience higher transaction volumes, with more competition for resources from different applications

with different execution characteristics. Even more important, these applications come from different parts of the business with different business priorities. All these factors make the job of managing performance and workload more complex and more difficult to resolve with existing tools.

Dynamic warehousing needs a workload manager to maintain service levels for each unique business application. To do this, InfoSphere Warehouse has the ability to explicitly control the resources (e.g., CPU, disk) an application can receive. By controlling the resources any given application can consume, in either a reactive or a pro-active manner, this solution renders the performance and service level of the applications predictable and manageable.

InfoSphere Workload Management lets you manage, enforce, and monitor workloads and their service levels for users across your warehouse:

- Enables consolidation of users into one warehouse
  - » Manage workload based on user priorities
  - » Prevent proliferation of data marts
  - » Prevent runaway workloads
- Empowers the convergence and management of transaction and query workloads onto one platform
  - » Real-time warehousing
  - » Operational data stores
  - » Batch workloads
- Improves monitoring and performance management
  - » Real-time monitoring
  - » Real-time resource management

An Integrated Approach to Building and Managing Trusted Information

IBM has assembled a portfolio specifically designed to help organizations deal with the challenges of building and managing the information infrastructure. This portfolio, called *InfoSphere*, accelerates the delivery of trusted information throughout an organization. The portfolio accelerates client value and reduces risk in critical information projects. We have now discussed the key components of the InfoSphere portfolio (Figure 3.5):

- InfoSphere Information Server
- InfoSphere Master Data Management
- IBM Industry Models
- IBM InfoSphere Warehouse

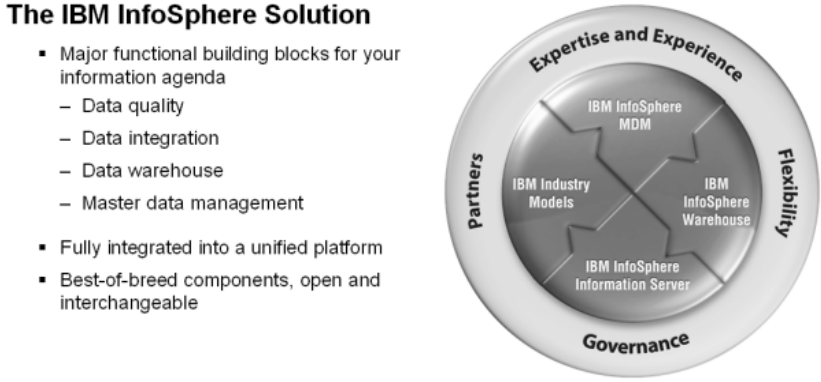

Figure 3.5: The IBM InfoSphere portfolio

InfoSphere Information Server specializes in integrating data across a heterogeneous landscape and delivering complete and accurate information when and where it is needed. A common target of this data is InfoSphere Master Data Management, which manages a master view of key data elements, such as customer, product, account, and location, over time. InfoSphere Warehouse provides a foundation for enormously scalable data warehouses, with embedded analytics and information lifecycle management capabilities to manage, control, and optimize the data. And providing acceleration for all these pieces are the IBM Industry Models, which contain industry-centric domain knowledge to help organizations achieve better results faster.

Getting Started

An accelerator to understanding your data before building the data warehouse is *InfoSphere Foundation Tools*. This open toolset offers unique capabilities to help you "get your arms around" your data. The tools help you discover and understand the data you have, design your trusted information structure, and govern your information over time. Key functional requirements they address include

- Building industry context into your information
- Information modeling
- Understanding your information sources
- Understanding the definition of your business information
- Leveraging an information repository
- Using sample information accelerators

Now that we have reviewed the key information-based components necessary to build a reliable and trusted information infrastructure, it is time to analyze the business intelligence and performance management tools you can use to exploit that infrastructure.

4

Maximizing the Business Value of Information with Business Intelligence and Performance Management

We all know intrinsically that information is valuable only if it is put to good use. How is it, then, that most organizations have focused more on collecting, managing, and archiving their information? The good news is that a shift is now starting to happen. Organizations are realizing that to be more competitive in the marketplace, or to provide superior service when managing costs, they need to *use* this information that has been squirreled away — and its use will create competitive advantage.

This chapter begins with a definition of challenges organizations face when seeking to place information into business hands to drive decisions that impact the business. We introduce the concepts of business intelligence (BI) and performance management and discusses the typical capabilities needed to drive the full decision-making process. We then move on to discuss the different business user requirements for access and interaction with information based on roles and tasks. We conclude the user-centered discussion with a look at how organizations need to be aligned across, and up and down the organization, for effective, collaborative decision-making.

With all the business demands for BI and performance management comes the requirement to put in place the right foundation. This chapter goes on to discuss the platform characteristics required to enable IT to source, organize, and store information that fuels the decision-making process and ensure IT can cost-effectively deploy, scale, and respond to business change. We also look in detail at the IBM Cognos 8 platform, built from the ground up on a modern service-oriented architecture (SOA), central to the Cognos performance management system. This platform forms the foundation for delivery of capabilities to the broadest user community and provides central services for the information flow across the full Cognos portfolio to each and every user.

For context, the chapter provides a brief overview of the Cognos product portfolio together with a discussion of how to accelerate time-to-value with Blueprints and

Analytic Applications. We conclude with a suggested path for getting started on your journey to performance management — a journey, once embarked upon, that will ensure everyone in your organization, across business and IT, is more informed, more engaged, and more aligned to drive better business outcomes.

Business Challenges

Data can be your organization's greatest asset. Whether it comes from day-to-day operations, enterprise resource planning systems, data warehouses, operational data stores, or other sources, data is the key to understanding and managing your business performance. Used effectively, it tells you how you are doing relative to plan, why, and what you must do to succeed.

But reaping the benefits of data isn't easy. Data is typically fragmented, often incomplete, and not readily available in a form that can be used effectively by the people who need it for decision-making. And the users who want to access and interact with that information often face multiple disparate reporting and analyses tools, disconnected planning processes, and capabilities that either limit them or are technically too advanced for their use. To realize the benefits of data, IT must be able to give the business easy access to data from all sources, no matter where it resides, in a way the business users can understand, trust, and use. And, vice versa, business users must be able to contribute their plans, insight, and expertise to enrich the information and drive the business forward.

Many factors influence your company's performance, but few areas are more important than your organization's decision-making ability. Getting answers and acting on them means integrating reporting and analysis, planning, and measuring and monitoring — across the organization. This integrated approach is the hallmark of an organization that has expanded from simple understanding with reporting, to improved understanding and better decisions with business intelligence, to better decisions and accountable actions with a performance management system. The IT organization is expected to seamlessly integrate information, technology, and process. The wide range of required capabilities, the number of users and communities involved, the large and varied data sources, and the complexity of enterprise infrastructures make this no easy task. So platform and architecture are also important considerations for IT when assessing performance management solutions.

Answer Three Critical Questions to Drive Business Performance

Effective performance management is possible only when people throughout the enterprise have the information they need, in the way they need it, to make the better business decisions. Better decisions enable them to meet operational goals and thus their strategic objectives. To be effective, this information must not only be delivered in an informative and engaging way so business can align; it also must be timely and accurate, and it must present a commonly agreed-upon, consistent, enterprise-wide

view of reality across departments, divisions, and corporate functions. This job is especially challenging because many organizations must convert vast amounts of data from day-to-day operations, ERP systems, data warehouses, operational data stores, and other sources into a business-usable form that traverses these information silos to support the decision-making process.

The business result of building a solid foundation for performance management will be a system that delivers a common understanding and actionable information based on answers to three critical questions that determine the performance of your organization (Figure 4.1):

Figure 4.1: Three critical questions to answer about business performance

- *How are we doing?* — Measuring and monitoring performance using scorecards and dashboards lets you track your key metrics.
- *Why?* — Reporting and analysis let you see data, gain context, understand trends, and spot anomalies.
- *What should we be doing?* — Planning, budgeting, and forecasting let you set and share a reliable view of the future.

The right answers to these three questions can mean the difference between success and failure to organizations in every industry, whether commercial, public sector, or non-profit. This holds true for organizations of any size, from global Fortune 100, to large national organizations, to small mid-market. All organizations alike need the same ready access to information so they can plan, understand, and optimize their performance with minimal IT resources and assistance.

Break Down the Silos Between Organizations and People

The immediate need for access and interaction with information to drive the business knows no boundaries. In the past, to meet the immediate needs of users in different

departments across an organization, IT often resorted to addressing one-off requests. These projects were delivered by implementing departmental reporting, analysis, or other decision solutions, each with their own data to support them — creating unfortunate results.

The first result was the introduction of significant data reconciliation challenges and difficulty in determining which information was "right." Different analysis tools, point-reporting solutions within ERP, one-off dashboards, and spreadsheet-based products all added their own "views" on how to present the information. Business users lost trust because data didn't match and it was hard to put together a complete picture. And IT had to dedicate too many resources simply to manage the patchwork of data stores and solutions — limiting the size of user rollouts.

The second result was usability challenges — it was simply too hard for business users to obtain the information they needed. Using point solutions or home-grown reporting or dashboarding tools created a patchwork of user experiences that the business was unwilling to learn and adopt. And as IT struggled to deploy, support, and maintain disparate tools and technologies, complexity increased.

The third result was an inability to gain a complete picture of how the organization was performing and acting upon complete decision-making cycles. Information couldn't be shared easily across user communities because data was departmentally owned, and users couldn't follow decision paths from identifying an issue, to understanding its cause, to adjusting plans. Decisions were made in silos based on partial information rather than in the context of strategic goals. The world of decision-making was data-rich and information-poor, poorly adopted and siloed.

To overcome these challenges, organizations have now set about discovering how to engage the business user with information that is relevant and accessible, how to set and communicate strategic goals and tie operational decisions to them, and how to ensure the foundation they put in place has both the ability to address current needs and the flexibility to adapt quickly to change to keep pace with the business.

Deliver Information Interaction When, Where, and How It Is Needed

From simple report viewing to in-depth what-if scenario modeling, from casual user to financial analyst to executive, business intelligence and broader performance management solutions need to address all the information interaction needs of the organization to drive the business forward. For example, a dashboard can't answer why a sales metric is higher or lower than expected unless it is tied to a report that shows the detail of what sales make up that figure. And that report should be tied to interactive analysis, where business users can compare sales trends over time and between regions and products to understand what is causing those trends. And this understanding must fuel actions, plans, and adjustments to get sales back on track or to take the learning from one successful region to improve the performance of another.

An effective performance management implementation includes BI capabilities for reporting and analysis, dashboarding, and scorecarding. It also includes the planning, budgeting, and forecasting capabilities to take action based on the gained understanding. With BI and broader performance management, the organization can make better decisions and create accountable actions that move the business forward.

Reporting and Analysis

Most organizations start delivering information to the business with reporting and analysis to deliver an understanding of why the organization is performing well or poorly. The report provides details, and the analysis provides exploration to uncover trends and patterns. Together, these two components supply a seamless, guided experience or exploration from one number to the next, from one question to the next — without requiring the user to switch tools or data sets or to ask IT to produce another report.

With reporting and analysis, users have access to friendly information across all data sources so they can:

- Align the conclusions they reach on a shared, common version of the truth
- Access the information in context, find trends, and spot variations, risks, and opportunities
- Understand the reasons behind the organization's results and trends

Figure 4.2 summarizes the functions provided by reporting and analysis tools to assess the factors contributing to organizational performance.

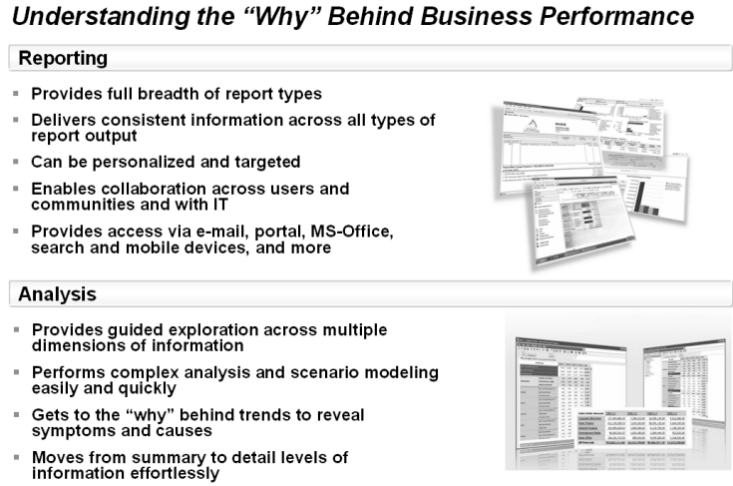

Figure 4.2: Reporting and analysis functions

Dashboarding and Scorecarding

Business managers, executives, and operations managers need to be able to monitor how the business is doing and know immediately where they should focus their attention to get the business back on track or to exploit an opportunity in an area where the business is performing well. A *dashboard* provides an at-a-glance view of all the critical key metrics that drive the business. A *scorecard* provides a map of how metrics inter-relate to each other and a comparison of status against a benchmark or goal to evaluate whether the business is on or off track. Like reporting and analysis, both dashboards and scorecards require the ability to seamless explore data to find the facts behind the numbers.

All too often, organizations deploy a dashboard that is not linked to the reporting and analysis environment — dead-ending the user with knowledge that a problem exists but with no ability to understand what is causing it. With dashboards and scorecards seamlessly linked to reports and analysis behind the figures, users can quickly focus on parts of the business that need attention so they can:

- Monitor performance against targets and hold people accountable
- Quickly identify the areas of the business that need their attention
- Uncover cause-and-effect relationships across metrics and departments

Figure 4.3 summarizes the functions that dashboards and scorecards provide for monitoring business performance.

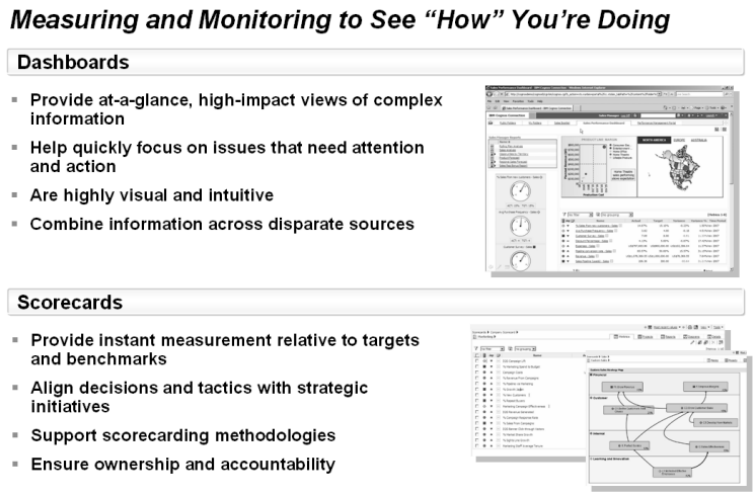

Figure 4.3: Dashboard and scorecard functions

Planning, Budgeting, and Forecasting

Finance and other department managers with cost or revenue generation responsibilities turn the results of analysis and decision-making into future plans, budgets, and forecasts. Armed with an informed understanding the "why" behind performance, these business users focus on what is needed to get the business back on track or to take advantage of new business opportunities. They "close-loop" the decision-making cycle with forward-looking action to answer the organization's question "What should we be doing?"

And, just as reporting and analysis should be tied to scorecarding and dashboarding, users require planning, budgeting, and forecasting to be a seamless next step in the decision-making cycle, so they can:

- Create plans and budgets that connect (unlike spreadsheet-based systems)
- Adapt plans organization-wide as business conditions change
- Engage departments outside of finance in the planning process for greater accuracy and collaboration

Figure 4.4 summarizes the functions provided by planning, budgeting, and forecasting for decision-making.

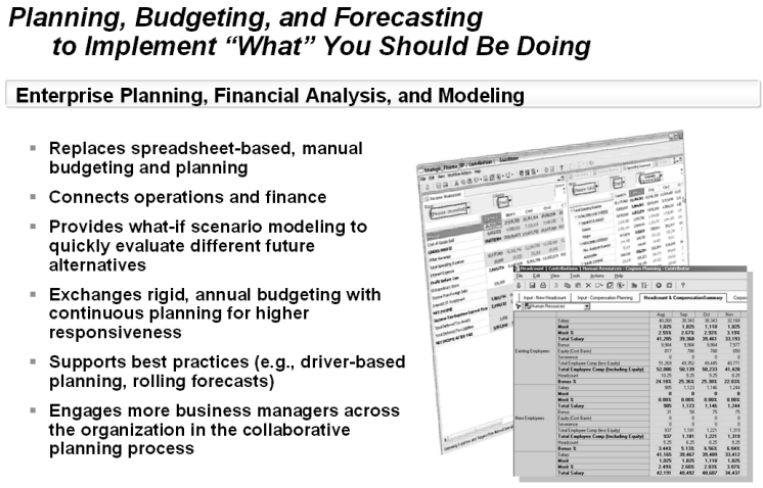

Figure 4.4: Planning, budgeting, and forecasting functions

Extending the Reach Beyond the Traditional Workplace

As we have mentioned, engaging all decision-makers in an organization is key to ensuring the broadest use of trusted information across the business. The easier the access to information, the more likely business is to adopt that information. As an example, consider e-mail adoption — think about how many e-mail messages users

answer at odd hours and on weekends now that e-mail is accessible on mobile devices day and night.

The key to reaching beyond the traditional workplace is ensuring that the information users see in reports, dashboards, and analysis on the Web and in the office is exactly the same as the information they receive on the mobile, in Microsoft Office, or as a result of enterprise search. Not only does this consistency instill confidence in the numbers, but it creates broader access without placing additional burden on IT to re-author information or analyses everywhere users want to be. Such flexible access provides numerous benefits, including those noted in Figure 4.5.

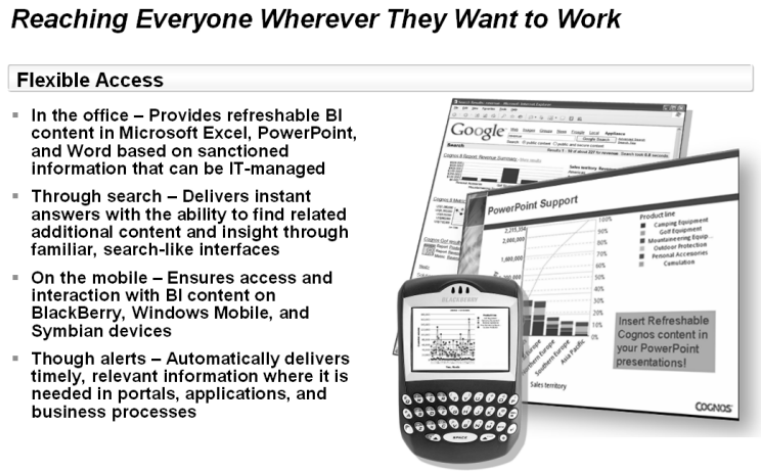

Figure 4.5: Benefits of flexible access

Extending BI to everyone's familiar workspace requires users to be able to:

- Use the office productivity tools they know while staying connected to the library of reports and analysis and trusted information source as they construct their own views
- Use free-form text search to find relevant reports on that same trusted information without requiring them to even know they are in BI
- Use their mobile device as a location where they interact with those same reports
- Receive alerts that go beyond basic notification to deliver relevant information in context and compress the time to action and resolution

Make Information Relevant and Fuel Collaboration

Traditionally, BI and planning tools have been built in a capability-centric way — what capabilities are required for reporting, what capabilities are required for

analysis, and so on. Creating information relevancy and fueling collaboration requires a change in this approach. This shift is from a focus on discrete capabilities to a focus on what combination of capabilities is required for a particular user role and the tasks, or decisions, users in that role want to perform. A review of the breadth of user roles for BI and performance management makes clear what is required for the decision-making process in an organization.

Target Information Interaction Relevant to Each Business User's Role

Most BI software targets the professional author or power user. This group accounts for a mere five percent of the employees in a typical organization. For BI to drive better business performance, organizations must engage the over 90 percent majority. This diverse group includes casual business users, managers, business and financial analysts, operational line managers, and executives. All these users need the right information and the right level of interaction for their task at hand.

Some users need dashboards to view "at-a-glance" status. For example, a sales manager might want to see where the sales force is with respect to quotas (who is on track, who is lagging behind, who is ahead), how much budget has been used (training, trade shows, travel and entertainment), or what new sales opportunities exist and who is driving them. A human resources manager might have an entirely different set of requirements for a dashboard. A CFO may need to track key metrics using a formal scorecarding methodology such as the Balanced Scorecard. A CIO might need powerful production reporting and ad hoc capabilities to track and understand IT resource allocation and use. Operational managers can't wait for daily reports but need a dashboard with up-to-the-minute information to monitor status and make near-real-time decisions about the allocation of resources (e.g., in a call center). A business analyst might need to construct four or five different pricing models and see their relative impact on profit and loss before making a final recommendation for forward planning. Finance users need to run reports against financials and may need to do in-depth analysis.

It requires discipline and a good understanding of the needs of different business users to recognize that often a "less is more" approach is prudent when it comes to functionality. Casual users are often overwhelmed with the rich functionality and interactivity provided by most solutions today. They are best served with easy access to regularly distributed reports and capabilities that use simple business terms and present information clearly for consumption. Business analysts, on the other hand, would be frustrated when faced with this simplicity. They want the freedom to explore complex hierarchies with ultimate flexibility to contrast their own views of information without restraint.

Yet, despite significant investments in information technology, most companies admit that employees are making decisions without the benefit of timely, accurate information. Why aren't more decision-makers leveraging business intelligence? Why aren't those decisions being tied into actions and future plans? Well, decision-makers

adopt BI and planning only when these solutions serve their need and level of interaction and when the information is trustworthy, timely, and relevant to the task or decision at hand. In short, the IT organization needs to deliver personalized business intelligence and planning, with personal views of information, without the cost of creating customized applications for each and every user.

We can divide the user communities in organizations into three broad groups:

- Users who want to consume and interact with the information — typically casual users, executives, business managers, partners, and suppliers
- Users who want the freedom to explore and create fresh perspectives — typically business analysts and financial analysts
- Users who create, format, and deliver content to others — typically professional report authors who reside either in the business as power users or in IT as part of a broader BI project team

Implementing a successful BI and performance management solution relies on engaging each of these communities and aligning them to work together to cover the complete decision-making process (Figure 4.6).

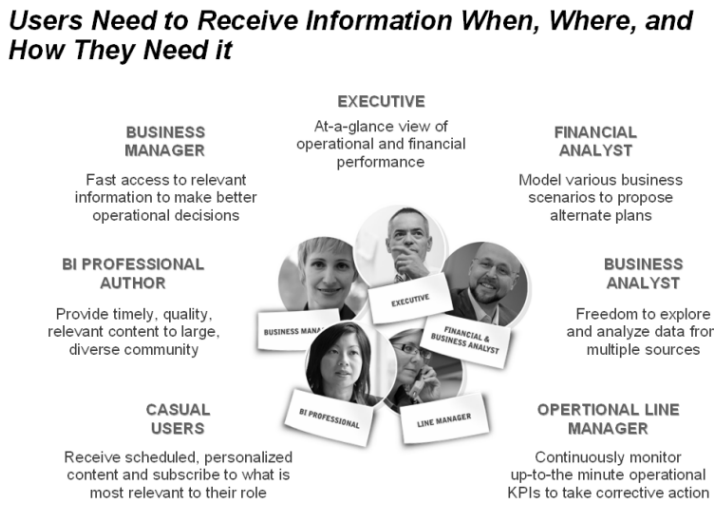

Figure 4.6: Targeting information interaction relevant to each user's role

Engage Users Who Want to Consume and Interact with Information

Users in the first group — those who want to consume and interact with information — want to receive information that is simple to understand and follow a guided path, when required, to additional information that might be useful to them.

Casual BI users represent the majority of potential users in an organization and account for the broadest number of users who access BI from a browser. This group includes sales people, operational workers, administrators, even customers and

suppliers accessing information via the extranet. What makes this group unique is that they need access from anywhere and with very little training. Casual users want:

- The ability to subscribe to regularly distributed reports
- Simple navigation
- Clear information presented the way they need to see it
- Secure, personalized, browser-only access

Executives, as organizational leaders, also require simplicity, much as casual BI users do. The difference is the breadth of information they see and the way that significant information from across the organization is highlighted for them. What makes this group unique is the need to tie operational metrics to more strategic perspectives. Executives want:

- At-a-glance awareness of organizational health
- Regular updates and status alerts about their specific areas of responsibility
- Secure access to business information from anywhere, including their mobile

Business managers represent approximately one quarter of the business users in an average organization. Users in this group have a wider range of responsibilities than casual BI users or executives and demand a wider range of self-serve access to BI capabilities. Business managers want:

- Drillable information that offers detail, context, and the ”why” behind the numbers
- Basic BI authoring that is easy to use, with integration to presentation and spreadsheet tools
- Occasional ad hoc queries, analysis, and metrics management

Operational line managers represent the user community who require information updated on a more frequent basis to make intra-day, intra-hour decisions. These users are similar to business managers in that they have a wide range of responsibility and demand a wide range of self-serve access. They are unique in their need to often see aggregate temporal views of streaming information sources. Operational line managers want:

- Time-critical views of operational key performance indicators (KPIs) based on near-real-time information
- Drillable operational data to make rapid, quickly implementable decisions
- Self-service to create personal views and occasional ad hoc queries
- User-defined thresholds, alerts, and actions to focus immediately on what is important

Enable Users Who Want the Freedom to Explore and Create Fresh Perspectives

The second group of users wants to understand the "why" behind the numbers and so require freedom to explore. These users need access to lots of information and a high degree of interaction with that information. They want to compare seemingly disparate information to find cause and effect and to build their own fresh views and perspectives of information not available to them from source systems or the data warehouse. Some also want to test different scenarios to determine a new or adjusted plan to present to the business manager or executive.

Business analysts represent a segment of the business manager population. They are unique because they are very savvy in both their analytical and technical abilities. Like professional authors, this group of users often creates content for others to consume. However, unlike BI professionals, the content they create contains new insight, a new proposal, or a new plan derived from insight they want to share about the business. Business analysts want:

- Freedom to explore and analyze data from many sources
- Professional report-authoring capabilities with a simplified interface to create statement-style reports
- The autonomy of spreadsheets without the burden of having to manage the data

Financial analysts require all the capabilities the business analysts require in their desire for freedom of exploration and use of spreadsheets. They differ from business analysts in that they are often tasked with determining new business approaches by evaluating alternative business scenarios. Financial analysts are more likely to want to see the financial and business impact of their models and then recommend a go-forward plan to business managers and executives for approval. In addition to what business analysts want, financial analysts want:

- What-if modeling to compare scenarios and gain fresh insight
- The ability to continuously reorganize information and use analysis to test hypotheses
- The ability to save scenarios and format a chosen scenario for presentation to the business
- A spreadsheet-like interface without any restriction

Ensure Users Who Create Content for Delivery to Others Can Easily Do So

The third group of users consists of the "go to" people who create content for others to consume. They may be adept analysts or technical users reporting to line of business. Or they may be technical people in IT who have a solid understanding of business requirements. Either way, they have deep knowledge of the availability and structure of the information, which they use to design reports, dashboards, and analyses in collaboration with users. They don't make business decisions themselves, but

they ensure that the users whom they serve get the information they need. There is often a backlog of requests for the output from this group of users, so enabling them to be more productive is very desirable.

Professional authors, also known as power users, are the traditional target of most BI software. These authors produce professional-quality content for others to consume. Professional authors want advanced BI capabilities:

- Easy distribution of a high volume of reports to many users
- The full range of BI elements and output types and ways to combine them into reports
- Complete control over the graphical look and layout
- Fast throughput performance for bursting reports to thousands of users and subsecond response for on-demand report access

Establish Alignment Across Organizations and User Communities

A complete performance management deployment spans many decision-makers, functions, departments, data systems, and business processes. Indeed, many decisions draw upon different data sources, involve multiple decision-makers, and dip into business process flows in an iterative way — creating a layer that sits above operational business processes and yet provides an iterative feedback loop into those core processes.

Implementing a collaborative decision-making process requires an approach that aligned tactical decisions with strategic goals and ties business processes together across organizational boundaries. This goal requires alignment in two directions:

- *Vertical alignment*, where strategic goals and objectives are set at the executive level, then tied to operational objectives at the functional level and, in turn, down to specific decision-making areas within those functions. The results and status are then bubbled up again.
- *Horizontal alignment*, where different functions can collaborate and execute decisions across business processes and departmental boundaries within the context of the strategy

In terms of business user collaboration, this vertical and horizontal alignment requires the context setting and active linking to business processes to ensure decisions and actions are implemented. It also requires that all users can share their reports, insights, and analyses with each other — not only to inform their colleagues but also to ensure they can build on each other's perspectives in a collaborative and iterative way. This is possible only with a common framework, a collaborative set of capabilities, and a solid foundation so each user isn't required to rework, create, or explain his or her own siloed view and reasoning behind decisions.

In terms of information requirements, this alignment requires every decision-maker be able to rely on a consistent and complete set of information, a method

to capture and share decisions and plans within the performance management system, and a shared, agreed-upon set of definitions — a common language that defines how the business want to shape the information to underpin its drive to business performance.

Putting the Right Foundation in Place for Performance Management

Optimizing business performance means getting answers and acting on them. This challenge requires integrating reporting, analysis, planning, measuring, and monitoring capabilities across, and up and down, your organization. It also requires seamless integration of people, information, technology, and processes. The business intelligence and performance management infrastructure must be able to deliver on these requirements and support the increasing number of user communities involved. It must also support the increasing breadth of interaction styles users require — from simple views and mobile access, to exploration and constructing/deconstructing aggregate views and plans, to sifting through volumes of data to present significant key metrics and casual relationships.

For users to adopt the solution and be able to collaborate based on a common view of information, the solution mustn't be fragmented, with different tools accessing potentially different data sets in different ways. A user should not have to switch from one tool when reviewing a dashboard to another when conducting analysis just because the underlying tools or data structures differ. And IT shouldn't have to deploy, manage, support, and maintain multiple disparate technologies, authoring, and metadata environments. Such scenarios often arise when a BI or performance management system is built as a patchwork of old technology components that attempt to give the appearance of complete solutions. An organization is most effective, from both an IT and a business perspective, if everyone in the business receives a common user experience, a common view of information in terms they understand, and the performance, scalability, and reliability they expect.

Having a solid foundation in place to address today's requirements as well as those in the future is critical to ensure sustainable benefit. With the right architecture in place, the solution can change and grow as the organization's needs evolve, all without adding an extra burden on IT.

An effective platform that supports the decision-making process of BI and performance management and can leverage and contribute to the trusted information in an organization must provide a way to:

- Optimize access to all data wherever it resides, inside or outside the organization
- Organize information in terms that business understands and capture business-entered performance drivers

- Ease deployment and ongoing management to meet and exceed service level agreements
- Fit with existing infrastructure choices and adapt to respond to future requirements

Cognos provides a performance management platform to address these requirements. Central to this foundation is the *Cognos 8* platform, built from the ground up as purpose-built, modern SOA. With these capabilities in hand, IT has the capacity to deliver on the business needs of today and grow the solution as the organization expands its deployment and use of performance management more broadly — all without putting an extra burden on IT resources.

The Cognos 8 platform is built using today's modern design for applications, grouping architectural components into three tiers, as depicted in Figure 4.7:

- A *presentation tier* that handles all user interaction in the Web environment
- An *application tier* with purpose-built services used to handle all processing
- A *data tier* supplies that access to the widest range of data sources and stores the performance drivers entered and managed by the business

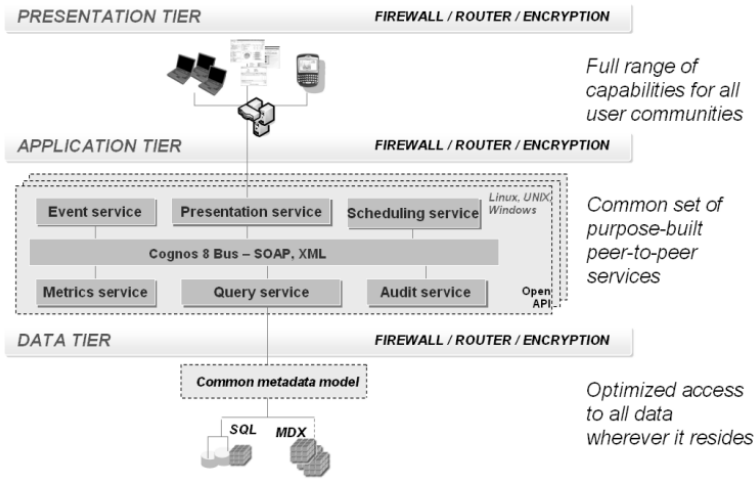

Figure 4.7: Cognos 8 architectural components

The separation of the architectural components into tiers supports the secure deployment strategies demanded by organizations whose data and infrastructure are secured and closely guarded by firewalls. This multi-tier approach also ensures that when a request is submitted to a Cognos 8 solution, the right processing happens at the right level and in an optimal way to serve the broad range of business users. Processing on the presentation tier, for example, can mesh with existing load-balancing routers, ensuring requests are distributed appropriately as they come in. Processing at

the application tier can be distributed across services and servers. Processing at the data tier can take advantage of the optimized database engines.

Optimize Access to All Data, Wherever It Resides, Inside or Outside the Organization

Virtually every organization has multiple different data assets, applications, and data integration technologies that store operational information and help the flow of transactions across those operational systems. Many organizations have started tackling this complexity and fragmentation to ready information for decision-making by implementing data warehousing or building data marts. These projects are a great start to making operational data available, reconciled and with high quality, to BI and performance management. Chapter 2 discussed the tasks and capabilities involved in creating this information layer.

Within the BI and performance management layer, the first step to delivering information to the organization for decision-making is to gain optimized access to that data regardless of where it resides. The key capabilities here are:

- Gaining complete access to all data, wherever it resides
- Flexibility to leverage any combination of sourcing strategy, now and over time
- Optimization for user response times by leveraging multidimensional storage and caching

The bottom line is that users need confidence that the data they want is going to be available to them when they want it. And IT needs the solution to provide breadth of access to all systems, with the flexibility to select the right sourcing approach and optimized storage to meet those user demands.

Gaining Complete Access to All Data Wherever It Resides

With Cognos, organizations gain consistent access to all data sources through a universal data access layer and single query service within the Cognos 8 platform. Data professionals access data using published interfaces across relational sources such as operational data stores and data warehouses; online analytical processing (OLAP) sources (e.g., Cognos PowerCubes, Cognos TM1, Hyperion Essbase, Microsoft SSAS, and SAP BW); transactional applications (e.g., PeopleSoft, Salesforce, SAP R/3, and Siebel); and modern and legacy sources (e.g., Cognos Now!, Lightweight Directory Access Protocol, mainframe systems, and XML).

Cognos provides a single query service that generates consistent and predictable queries to these interfaces. It understands and leverages source-system strengths by using a combination of open standards (e.g., SQL99), as well as native SQL and native MDX to optimize data retrieval from all these different data providers.

Flexibility to Leverage Any Combination of Sourcing Strategy, Now and Over Time

Once data is accessible, IT must determine the best sourcing strategy based on how the information is to be used by the business in the decision-making process. If the business is looking for historical summaries or a comparison of plan versus actual, the sourcing strategy must be to aggregate, calculate, and capture the history of the data. In other cases, the business may need immediate, up-to-date insight into information sourced directly from one operational source or from an on-the-fly view across two or more operational sources. Another consideration is that if a data warehouse is already in place with a standard data integration technology, the data professional will want to leverage what already exists.

The key requirement, therefore, is that IT needs freedom to choose the best method to source each data set, to mix and match sourcing approaches depending on need, and to have the flexibility to change approaches as business demands change or as new data sources or data views are required.

Cognos provides the full range of sourcing approaches:

- *Direct sourcing* to provide instant access to a source system or data warehouse
- *Virtual sourcing* to federate multiple sources on the fly to deliver a combined, high-performance view
- *Dimensional ETL sourcing* to conform data and manage complex hierarchies into relational formats optimized for reporting and analysis access — including creating data marts with a dimensional framework for the "last mile to BI" from enterprise data warehouse sources
- *Integration with ETL and data integration technologies* such as IBM InfoSphere DataStage or IBM InfoSphere Change Data Capture and other vendor's technologies to leverage existing integration technologies at the organization

Optimization for User Response Times by Leveraging Multidimensional Storage and Caching

Often, data sources — even existing data warehouses — are not optimized structurally to deliver high performance in response to complex or intense queries generated when complex reports are run or when in-depth analyses across large data sets are executed. And sometimes, these data sources do not contain all the information a user would like to combine in the shape and format to perform a particular type of analysis. In these situations, caching is required.

Cognos provides the full range of caching approaches:

- *Cognos Transformer* provides the ability to easily construct, build, and publish Cognos PowerCubes (OLAP data cubes) as caches of data for exploration and analysis by large communities of business users. Transformer converts high volumes of data into hierarchical, compact cubes

that deliver high performance and predictable response times. These cubes are accessible by any of the user capabilities, including analysis, reporting, mobile, and more.

- *Virtual caching* reduces load on source systems and yet gives IT the flexibility to construct mixed disk, memory, and database caching to optimize performance for user access without re-engineering back-end data sources.
- *OLAP caching*, in which queries are stored in server caches and new queries can re-use that cache for fast query response times on frequent queries, improves response times without requiring additional physical data storage.

Organize Information in Terms Business Understands and Capture Performance Drivers from the Business

For collaborative decisions to happen, everyone in the organization must be working from a common set of information that reflects the way they want to view and manage the business. The platform also must provide a way to capture performance drivers from the business, such as metric benchmarks and plans against which the business will align and measure performance.

For the business community to use the information, they must have confidence — confidence that reports, charts, dashboards, and metrics are all based on common data regardless of the data sources involved, what country they are in, or how they access the information, whether on the Web, in an office, on a mobile device, or through search.

The information also must be delivered in terms business users can understand and use without any requirement to know SQL or to be technically savvy — and in a way that insulates them from the underlying physical complexity of the data sources. It must be multilingual so that data need not be duplicated for different countries or user communities because such duplication introduces the risk of errors. And it must be published so that each user receives only the information he or she needs — no more and no less, so the shared view is easily consumable. In this way, IT is freed from having to explain the data, duplicate it for different views or languages, or create all the versions of reports users want because the information simply isn't in a form they can use and trust.

Within the BI and performance management layer, organizing information delivery to the user community and capturing the performance drivers entered and managed by the business requires key capabilities that include:

- Delivering consistent information in terms the business understands and trusts
- Reducing the time and effort to build, manage, and support information models for the business
- Capturing and managing the viewpoint and performance drivers from the business

Establishing the organization of information in terms business understands and enabling IT to model information for the business in a flexible and sustainable fashion is the first foundation of information. Then, making sure users are engaged to shape their views and perspectives on the business and capturing their metrics, plans, and analysis scenarios ensure maximum use of the information in a two-way dialogue that drives higher use and, therefore, higher business performance.

Delivering Consistent Information in Terms Business Understands and Trusts

For information to be in terms business understands and can use, data modelers need to translate the physical source, massage the information, and make it digestible. And they must consistently deliver that information to all business users so that everyone works from the same information yet is not overwhelmed by receiving more than what is relevant to them. To respond to the business, the data modelers must do all this efficiently so that information is readily made available before those business users decide to create information sets for themselves in spreadsheets.

Cognos provides the key capabilities to deliver:

- A *common business model* as a foundation for common understanding across the full range of capabilities, regardless of location, language, or information requirement. Textual descriptions can be provided for complex calculations and derived information. And information can be grouped into logical packages that are digestible by each user community.
- A *multilingual model* to capture text properties associated with metadata, such as descriptions and tooltips, so they can be translated, or dynamically substituted, upon access. This capability enables one model to support multiple languages, avoiding unnecessary duplication, permitting easy management of translated metadata in a single metadata location, and reducing report proliferation simply to support multiple geographies.
- The *ability to construct multidimensional PowerCubes* in quick response to various business demands for specific views of data and to do so without compromising the enterprise model or making the user wait until it is incorporated back in the warehouse. Without any requirement for database skills, business users, data analysts, and IT modelers can model complex dimensional structures, incorporate third-party data (e.g., SAP BW), apply business rules, and set security for business access. The PowerCubes are published as part of the enterprise model so business can use the full range of capabilities against them and be sure the information stays within the IT-managed environment.
- The *ability to publish sections of the common business model* to different users communities so they receive only relevant information, making it easier for them to find the information they want. This “model once, package for many” facility ensures users receives consistency, and it dramatically reduces

the IT time and effort required to deliver different models to different user groups.

- *Information security* to make the right information accessible to the right user group via granular row and column security and to limit users, groups, and/or roles to accessing only certain parts of the published model or PowerCubes. This capability leverages the existing security infrastructure and again reduces the need to proliferate reports because reports can be built more generically to contain all necessary information, with the certainty that when a user accesses a report, that user's view will be restricted to only the information he or she is permitted to see.

Reducing the Time and Effort to Build, Manage, and Support Models for the Business

Data modelers are responsible for managing the process of building the metadata model to deliver the highest-quality model to represent the business. In larger organizations, different modelers need to be able to collaborate in a virtual team, and each modeler needs to learn and leverage best practices. Creating a high-quality model from the initial design results in less troubleshooting and a more maintainable model that can absorb change along the way.

Cognos provides key capabilities for data modelers by delivering:

- *Model advisor*, an automated tool to identify areas of the model requiring further examination and recommend possible actions to improve overall model design. The advisor gives novice modelers a great tool to assist them and gives more experienced modelers a diagnostic tool to speed design. The results: faster development, easier maintenance, and better model performance.
- *Multi-participant modeling*, in which teams of modelers can access and work on different segments of the model and bring the segments together again for a common model. In this way, model conflict resolution can be identified early on in the project, and model consistency increases across the organization.
- *Multi-tier modeling* lets IT deliver information in a format that is easy for business to use and understand while insulating users from the physical complexity hidden underneath. This multi-tiered approach gives IT the flexibility and independence to evolve the underlying data sources, sourcing approach, and caching strategy without affecting the view provided to the business. This level of abstraction eases maintainability and ensures IT can absorb changes in data sources and infrastructure with minimal business impact.
- *Impact analysis and lineage* ensure IT can both understand the impact of data changes on existing reports and, in reverse, easily see which reports rely on what data and how that data traveled from the source. With impact analysis,

data modelers can understand the impact of model changes on users and create a work list to ensure all required modifications are estimated and completed. With lineage, IT can respond quickly to business questions about where the data came from and so increase understanding and data confidence.

- *Versioning* to enable publication of multiple versions of a metadata model to minimize impact on production while letting IT distribute new requirements to user communities.
- *Bidirectional third-party metadata integration* to reuse metadata that exists in data integration tools, metadata modeling tools, and source catalogs. This integration saves time and effort and increases model accuracy by eliminating retyping. It also provides for complete metadata lineage visibility into third-party metadata repositories for full traceability back to the source.

Capturing and Managing the Viewpoint and Performance Drivers from the Business

Many organizations think of BI as a view-only world in which information is delivered to the user in a single direction, from source and warehouse to report. However, to drive performance management, organizations must capture information from the user and from the decision-making process itself, including targets, plans, budgets, business hierarchies, analysis scenarios, and more. It is this business-entered data that defines how business users want to construct their view of the world of information, how they want to measure status against goals, and what new plans they want to put in place to shape forward behavior to achieve future business performance.

Cognos provides key capabilities that include:

- *Capture of metrics, annotations, and document attachments*. Metrics define what and how the business will be measured in terms of benchmarks and thresholds. Annotations and document attachments provide a way to capture commentary from the business user and associated articles for use in governance or to provide further background to a decision or plan.
- *Hierarchy, attribute, and dimension management* so the business can construct, govern, and share perspectives on how information should be organized and represented to support the decision-making process. Subject-matter experts in the business frequently maintain spreadsheets because operational systems do not capture this information. This is often because the operational systems lag the way the organization wants to manage the business and the transactional model can't be extended to store the extra attributes required. Or it may be because the business requires future structures for planning purposes that are irrelevant to the transactional systems until approved. By engaging business users to define and maintain the way they want to view the information, IT can increase user adoption and maintain governance, and the organization truly can manage to a performance view of the business.

- *Content store* to capture and manage the metadata of the solution itself — including reports, metadata packages, configuration information, and user and group preferences. This store serves as an indexed and searchable repository of information capital captured across the business.
- *Storage of what-if scenario results and plan versions* to capture and manage the analysis process and final approved plans that define how the organization is going to be managed and measured going forward.

This capturing of business viewpoints and performance drivers, together with the BI content of saved reports, charts, and so on, into a repository of business information for performance management is critical to putting decisions into action. It ensures that everyone is informed and is working from the agreed common viewpoint, with a common plan to drive achievement of operational goals against corporate strategy.

Ease Deployment and Ongoing Management to Meet and Exceed Service Level Agreements

Technology hurdles such as users without access, poor performance, and random system errors shake business confidence. IT needs to trust the technology to enable them to meet required service level commitments, give users the right access, and minimize the risk, extra effort, and overtime for the IT team. This confidence needs to be there at the time of initial deployment and should grow as the solution extends across the business.

IT managers and solution and security administrators require the BI and performance management solution to:

- Deploy easily and perform predictably in response to workload
- Provide system management capabilities to meet service commitments
- Ease the process of introducing change

So, to maximize adoption by the business community, IT requires ease of implementation, administration, and management to roll out to large user communities with confidence that the solution will perform.

Deploy Easily and Perform Predictably in Response to Workload

From the initial deployment to ongoing rollout across broader user communities and information sets, IT requires a solution that is predictable in behavior and provides the level of security that will stand up to today's scrutiny.

Cognos provides key capabilities that include:

- *Self-aware peer-to-peer services* can self-start, self-spawn, and self-manage and are distributable in any configuration to suit an organization's environment for full fault tolerance with no concept of a master service or

bottleneck. The services are loosely coupled, so no service needs to know how the other services do their jobs nor what they are doing at any particular time. This capability provides tremendous reliability because redundancy is built in to ensure uninterrupted access. If any service or server happens to fail, other services can absorb the workload while the service tries to restart itself.

- *Linear scalability* provides the ability to scale up and/or out so IT can manage with predictable response times to workloads. It also means IT can predict future system requirements to accommodate broader rollouts.
- A *zero-footprint Web interface* means that for most users who access information via the Web, IT faces no client-side installs, downloads, plug-ins, or the need to change user PC settings or configurations. Solutions can be deployed to large number of users within minimal effort and can be upgraded without the need to consider PC environments.
- *Comprehensive security* ensures that the information available in BI and performance management is secure. The Cognos 8 platform has been tested and has passed an independent and detailed security evaluation conducted by Symantec. This security includes authentication with leading third-party security providers, authorization to capabilities and user rights, data-level access rights to all objects and information, an application firewall to prevent unauthorized access to services, and encryption of all information from source to user using industry-standard encryption algorithms, such as Triple Data Encryption Standard (DES) and Advanced Encryption Standard (AES).
- *Adjustable auditing levels* provide comprehensive auditing across services and centralize the audit logs, for IT to respond to audits or governance requirements. In addition, charge-back environments can receive information about who is hitting the system, when, and for how long.

Provide System Management Capabilities to Meet Service Commitments

To meet service level commitments, IT requires system administrators to have visibility into system processes and activity and to have capabilities that let them proactively manage the deployment more effectively.

Cognos provides key capabilities that include:

- *Task-oriented system monitoring*, so administrators have a consolidated view of the solution activity organized around administrative task groupings. This support provides a holistic view of what is happening in the system — ranging from report activity to specific dispatcher workloads. It also enables IT to allocate administrative roles based on tasks to avoid providing every administrator with complete access.
- *Proactive administration* to access all the metrics and summary views needed to troubleshoot or identify potential issues. With the ability to set thresholds

on hundreds of system metrics, the system can notify administrators of a potential issue before business users are affected or start to call for support. This feature also enables IT to identify system trends over time to optimize future operations.

- *The ability to take action on behalf of users for scheduled reports* lets administrators set run priorities and change those priorities at any time. With report re-run, they can resubmit jobs to re-run only failed reports while maintaining all the user credentials and report parameters. This capability increases business satisfaction and minimizes business disruption when issues arise.
- *Integration with third-party enterprise management systems* exposes system metrics to enterprise management system solutions (e.g., IBM Tivoli) to enable IT to manage the BI and performance management solution as part of an overall administration environment.

Ease the Process of Introducing Change

Once the system is up and running, it is rarely static. There are constant changes to address. With success and business adoption, users start requesting new capabilities that are available only in new versions of the product; they want to know when these capabilities will be available to them to use. In addition, IT may be assessing different hardware options or seeking to take advantage of the latest version of its database standard.

Cognos provides key capabilities that include:

- *Visual upgrade manager*, giving administrators the ability to test, compare, and record report results between two versions of Cognos to quickly identify and quantify any issues and so streamline the upgrade process. This capability reduces the time required to upgrade large volumes of reports to newer versions.
- *Automated environment validation* lets administrators test, compare, and record report results, making it easier to validate the impact of making changes to the environment in which the solution resides. This feature reduces the time required to roll out production changes.

Fit with Existing Infrastructure Choices and Adapt to Respond to Future Requirements

Regardless of the purpose of any new IT application and its fit for purpose in the business, IT managers must ensure important technology requirements and architecture considerations are addressed. They need to know that the technology supports the chosen enterprise standards of the organization, such as portals, security, operating systems, and more. They also evaluate the fit with their preferred deployment

strategy, influencing how and where an application is deployed and what hardware is required. Last, the solution must fit the organization's architectural blueprint and be flexible enough to adapt to future requirements and standards as IT chooses to evolve.

For the BI and performance management solution to fit with existing infrastructure choices and adapt to respond to future requirements, the requirements are:

- Leverage preferred technology standards today, and as they evolve
- Deploy to take advantage of today's infrastructure, and adapt to the future
- Extend existing capabilities with new innovations, and extend reach with new integrations

With the Cognos 8 platform, built from the ground up on a modern, service-oriented architecture, IT can achieve the ease of implementation and integration promised in an SOA world.

Leverage Preferred Technology Standards Today, and As They Evolve

The Cognos 8 architecture is open and standards-based, using Simple Object Access Protocol (SOAP), Web Service Description Language (WSDL), and Web Services for Remote Portlets (WSRP), among other standards, to ensure IT can leverage the technology in place today and provide the encapsulation to adapt to changes in the environment as the IT infrastructure evolves.

The Cognos architecture is:

- *Standards-based* — With open standards, the Cognos 8 platform gives organizations the flexibility to leverage existing IT environment standards without duplicating functions (e.g., security) or dealing with issues associated with proprietary application servers or rich client tools. Cognos is open to run on any Web server and complies with existing routers and firewalls.
- *Environment-neutral* — The Cognos 8 platform is not tied to a particular technology stack, and it can leverage any combination of existing environment standards, even mixing and matching within a single deployment. In this way, IT can leverage existing heterogeneous platforms, security, portal, and Web environments and can evolve those deployments and choices over time with confidence.

Deploy to Take Advantage of Today's Infrastructure, and Adapt to the Future

IT managers need the freedom and flexibility to deploy performance management where it best fits their infrastructure, now and in the future. For some organizations, this means large, powerful servers managed centrally. For others, the preferred approach is a distributed, cluster hardware environment. Still others may be adopting technology options such as virtualization. These choices are made to optimize the hardware environment and reduce the cost of operating IT systems. Therefore, a

performance management solution must be able to adapt to the chosen IT strategy to minimize the cost and effort to deploy.

The Cognos architecture adheres to:

- *Location transparency* — The Cognos 8 platform can be deployed as services across the network using the standard Internet protocols of SOAP and XML. As a result, services do not need to know where other services reside to communicate with each other. This support gives IT ultimate flexibility and a way to leverage available capacity across the enterprise network.
- *Peer-to-peer services* — Because the Cognos 8 platform is truly a service-oriented architecture based on peer-to-peer services and Internet protocol, IT can choose to deploy services as co-located or co-resident with other applications or databases to minimize traffic and maximize server usage. Each of these services performs a particular function and is loosely coupled in that no service needs to know how others do their jobs, what they are doing at any time, or where they reside. As a result, IT has infinite flexibility for services distribution and can deploy this solution to fit the broader deployment strategy. These advantages are especially important when managing peak usage periods, during which IT can simply leverage idle capacity to meet the increased demand.

Extend Existing Capabilities with New Innovations, and Extend Reach with New Integrations

When selecting a solution, it is important to consider how easy it is to extend that solution with additional modules to address new or unique business challenges or perhaps to integrate with other applications and technologies in the IT infrastructure, such as business process management, enterprise search, or transactional systems. This consideration is important not only for how your organization can extend the solution but also as an indication of how easy it is for the solution provider to add innovation that you can exploit in future versions.

The Cognos architecture provides the following features in this regard:

- A *fully documented and open API* to all platform services provides a simple way to integrate and extend capabilities using any integrated development environment (IDE). This support ensures end-user organizations and partners can extend functionality, integrate other solutions, or add business functionality to address specific user requirements.
- *Purpose-built services* provide the basis for augmenting existing functionality or adding a service while drawing on existing services, such as query or presentation services. Once the new service is introduced, it is available to all services in the platform. This capability minimizes the effort involved to innovate additional value because existing services can be reused, and changes don't have to be made in multiple places for all platform capabilities to take advantage of them.

With an open API and purpose-built services, the focus can be on innovation. For example, a new data source such as Salesforce.com can be added and made immediately available to all user capabilities. Or a new user access method, such as access via a mobile device, can be added and immediately all existing authored reports are available without requiring any re-authoring or changes to existing capabilities. In this way, new and unique business requirements can be addressed to enable IT to extend the way performance management can be delivered across the organization.

A Brief Overview of the Cognos Business Intelligence and Planning Capabilities

Figure 4.8 lists the business intelligence and planning capabilities available in the Cognos portfolio. The portfolio is centered on *Cognos 8 BI*, which delivers the broadest range of capabilities to the broadest community of users and is underpinned by the Cognos 8 platform to orchestrate the flow of information across the portfolio to enable the complete decision-making process.

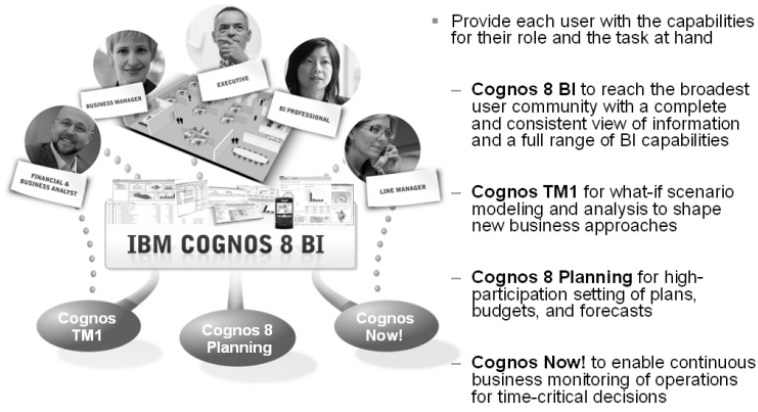

Figure 4.8: Cognos BI and planning capabilities

To speed time-to-value and to capture Cognos expertise in performance management, Cognos also delivers performance blueprints and analytic applications — all designed to accelerate deployment and ensure success.

IBM Cognos Performance Blueprints are predefined data, process, and policy models that help organizations speed their software deployments and drive faster return on investment. The Blueprints has been developed in collaboration with top-performing customers and performance management experts. They address several essential functional process areas as well as the unique needs of specific industries. In addition, Blueprints can be linked together, letting companies establish

dynamic connections that keep strategic objectives, operational plans, people, and activities aligned as business conditions change.

IBM Cognos Performance Blueprints enable companies to:

- Coordinate financial and operational planning
- Improve forecast reliability and make better resource allocations
- Quickly deploy an integrated, dynamic planning process
- Streamline project implementation schedules and reduce risk
- Reduce the total cost of ownership

The Blueprints include the following components:

- *Application Briefs* describe how the planning models for each Blueprint work.
- *Business Value Guides* describe the value opportunity enabled by the Blueprints and illustrate best-practice process and information flow.
- *Webinars* and *Web Demos* highlight a particular Blueprint in detail.
- *Data Models* are the actual predefined data, process, and policy software models for use with IBM Cognos 8 Planning.

*IBM Cognos 8 Analytic Application*s (depicted in Figure 4.9) are packaged, adaptable software solutions to help improve business performance by providing insight from the information locked in enterprise resource planning and other data sources. They include configurable connectors to popular ERPs, a data warehouse with a dimensional data model for performing analysis, and interactive packaged reports.

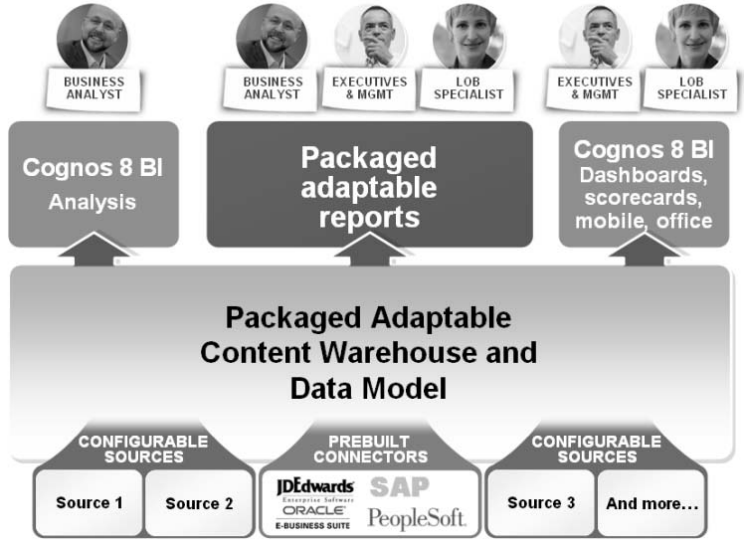

Figure 4.9: IBM Cognos 8 Analytic Applications

These solutions are faster to implement than building a custom solution, and they easily adapt to changing requirements in minutes, not months, dramatically reducing the total cost of ownership. Built on the Cognos 8 platform, they enable organizations to implement a solution quickly and to extend it easily using the full range of Cognos 8 capabilities.

IBM Cognos 8 Workforce Performance quickly provides insight into key areas of the workforce, such as employee strength, acquisition, development, retention, and compensation. It is available with connectors to Oracle, PeopleSoft, and SAP Human Resource Management Systems. IBM will introduce more IBM Cognos 8 Analytic Applications in other horizontal areas, such as financial performance analytics, customer analytics, and others.

Getting Started on Your Journey to Performance Management

We have touched on many topics in this chapter, from the approach business must take to turn the decision-making and planning process into better organization performance to the value of information optimization and organization centered on an SOA foundation to enable IT to be responsive to the business. For those organizations that are currently defining their requirements and reference architecture for BI and performance management, or for those seeking opportunities to improve existing solutions, we recommend focusing on three key areas on your journey.

1. Define your strategy and roadmap.

- Define the metrics and information "sweet spots" your organization requires to plan, understand, and optimize future business performance
- Ascertain by user community the set of capabilities each business user needs in order to address the three critical questions that drive business performance:
  - » How are we doing?
  - » Why?
  - » What should we be doing?
- Identify where to start by addressing today's most critical requirement, which for many organization starts with reporting to deliver information for use by more decision-makers across the organization.
- Set the roadmap for your organization's journey to broader performance management by recognizing the high-impact "next steps" of rollout that ensure value at every step — whether that be adding user communities, fueling additional business processes, extending breadth of capabilities, or expanding reach with new paradigms of mobile and search.

2. Set your solution requirements.

- Evaluate the breadth of capabilities provided and how the solution supports a collaborative decision-making process across business users and between business users and IT. Assign value to capabilities important to your organization, be they self-service for the business; access from anywhere at any time without switching tools or seeing different data; multilingual capabilities without re-authoring; a seamless experience for business users to navigate, explore, and capture decisions without requiring IT to support multiple environments; or other possibilities.
- Assess the solution's ability to effectively access and manage the flow of information to business users and, in reverse, to capture and manage the metrics, plans, and decisions the business defines to drive business performance. Assign value to the degree to which business users have a complete, consistent view of information in terms they understand and the extent to which they can define how they want to manage the business for performance without having to integrate, synchronize, and constantly worry whether the data is in agreement and governable.
- Assign value to how easily IT can deploy and manage the solution, not only for the first rollout but also when scaling in production. Evaluate how easy it is to troubleshoot problems and resolve them quickly, before the business is impacted, and assess your confidence in being able to meet and exceed service level agreements.
- Look under the covers at how the solution was architected because this is really the determining factor for your ability to adapt and evolve the solution in the future as requirements change and your IT standards advance. A modern SOA goes beyond claims of Web services interfaces to a truly purpose-built, peer-to-peer, services-based architecture to give IT flexibility, responsiveness, and the freedom to innovate new ways to deliver the value of BI and performance management across the organization.

3. Apply performance management principles to your own deployment.

- Define the metrics of success for your solution. Look beyond the go-live date, and focus on your coverage of decision-making processes, the breadth of information delivered and managed, and the rate of user adoption.
- Inform the business with scorecards and status reports that measure how the solution is delivering value and provide results as fuel for future investment in new areas.
- Look for ways to continuously optimize your solution. This may not always be to deliver more to more users. It may mean targeting improvements to the relevance of capabilities and quality of information to address specific areas of the business.

- Foster a partnership between business and IT to acknowledge delivered value and focus future investment. Engage in a continuous dialogue about how to optimize the solution to create and sustain competitive advantage.

5

Innovation with Trusted Information

How can information be the cornerstone to innovation? It is through information that companies can identify customer behaviors based on buying patterns, demographics, and psychographics, aiding customer segmentation and the development of targeted offerings. It is with information that organizations can gain visibility into the key performance metrics of the business using corporate dashboards. Information is how organizations document and optimize key business processes, and it is what satisfies the growing set of regulations with which every organization must comply.

Organizations need to establish an accurate, trusted view of information across business processes and applications to drive more consistent information throughout the enterprise. IBM InfoSphere software provides the foundation to support Cognos business intelligence and performance management applications, delivering a solution for trusted business information (Figure 5.1).

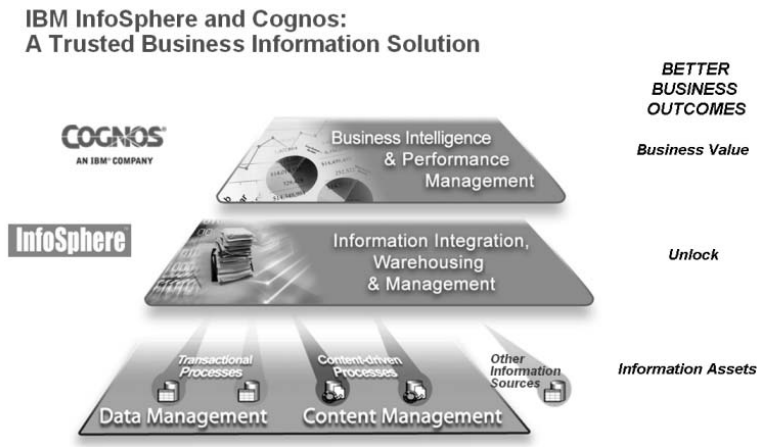

Figure 5.1: The IBM InfoSphere and Cognos solution

In this book, we have reviewed a number of technologies and the value they can provide to the organization. We have discussed the requirements for a robust, integrated information architecture in light of the demands for real-time data access to trusted information and user consolidation.

Such an architecture requires a common information framework that defines common terms and models for how information within a business relates across business systems. IBM has brought together the various capabilities to provide an integrated information solution that integrates, reconciles, manages, and analyzes data and content regardless of type, volume, or complexity (Figure 5.2). The sources can be different silos or different vendors' systems, and the data can be delivered in context, as quickly as required, to users, applications, and business processes. With this integrated information infrastructure, organizations can unlock the business value of their information for competitive advantage.

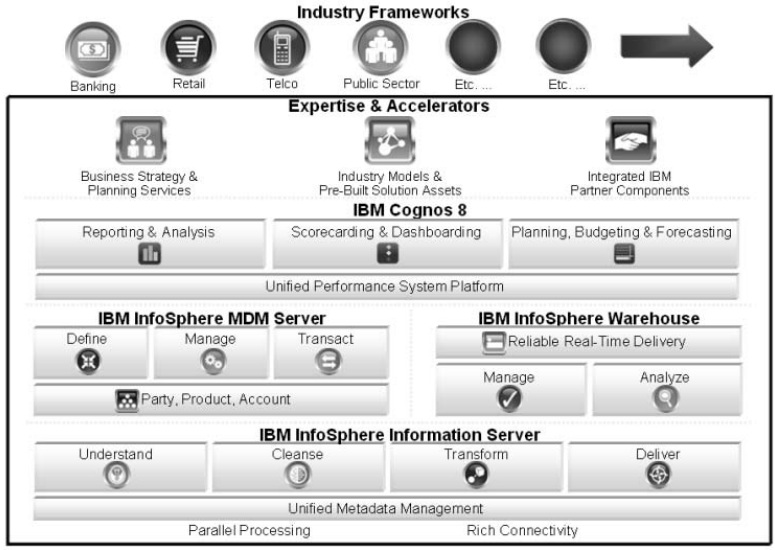

Figure 5.2: IBM integrated information infrastructure

Sustaining Success with Trusted Information

Business optimization is built on a solid foundation of information. These foundational information systems are not just a random collection of technologies but a combination of automation and best practices according to a design that enables a management-mandated implementation of business goals and objectives. These systems span multiple organizational functions and require both management concurrence and the integration of established and current systems, which can represent high volumes of data with a great variety of data formats. Deploying a business intelligence,

performance management, or data warehouse today is a task that cannot be trivialized and relies heavily on the quality and availability of data to populate the system.

Many companies have failed in their pursuit of innovation under the misguided thought that they could address the problem with the hottest items in technology. The reality is that information systems must support the strategic objectives of the organization. They can give insight into opportunities, provide information to aid in achieving those opportunities, and supply the ability to measure progress in attaining those objectives. Aligning with the goals, objectives, and initiatives of the organization can make the difference between success and failure.

Even those organizations that succeed in their initial efforts to establish trusted information warehouses face challenges in making the process sustainable. One of the most important aspects in guaranteeing the viability of these systems is to measure the business impact of such investments and understand the pressures of continuous business growth. One should not underestimate the effort of translating such requirements into IT solutions.

Other organizations fall into the trap of complacency, treating their data warehousing systems as standard applications that require neither governance nor constant assessment. But complex systems such as business intelligence, performance management, and data warehousing systems require a different kind of attention from transactional systems such as enterprise resource planning applications, for which the measures of quality of service (QoS) and growth are simpler and deterministic.

Measure Business Growth

The information agenda is built on a solid approach that integrates strategy, information governance, and enterprise information infrastructure with an inclusive implementation roadmap. It is a comprehensive, enterprise-wide plan that the CIO, working with line-of-business colleagues, can use to achieve strategic initiatives of the business.

A key aspect of the effort to develop such a plan is defining, prioritizing, and measuring business results. Although many companies can succeed in the short term with narrow definitions of such results, a more comprehensive and long-term vision allows for a sustainable process to maintain the necessary focus on growth and enhancements in the information warehouse.

The goal is to transform the organization and unlock the business value of information across data and content silos. The information strategy provides an end-to-end vision for all components of the information agenda and is driven by an organization's business strategy and operating framework. With the information agenda, organizations can deliver trusted, dynamic, enterprise information to optimize business performance. With measured accomplishments, you can:

- *Socialize the successes*. To gather support for governance, investment, and cross line-of-business collaboration, make sure other lines of business understand the value of existing information warehouse results.

- *Reinvest part of your information warehouse profits.* By defining an information agenda roadmap, you can organize projects to fulfill the plan. Each project should have a set of measurable benefits, such as cost reduction, an increase in strategic position in the market, or an increase in bottom-line revenue. The results of initial projects can thus help finance subsequent projects in the roadmap, either by reducing the size of the investment due to reuse of existing capabilities or by reinvesting earlier savings or profits to finance new innovations.
- *Create an innovations incubator to drive new businesses*. Don't wait for requirements to define what the next big area of information innovation will be. Treat innovation in your information systems as a research investment, using entities such as centers of excellence and business user groups to help guide future needs in information architecture.

Plan for Capacity and Business Continuity

Many organizations struggle with justification to invest in reliance and QoS for information systems that are not directly tied to business operations. Although many feel these systems should be highly available, the justification and business case are not always tangible, especially when budgets are tight.

When data warehousing projects are successful, business users, processes, and decision-makers become so dependent on these systems that they should be considered mission-critical. The best way to anticipate capacity and ensure availability in these information systems is to define the impact of *not* having your information available. Balance your investments in system resilience against the risk of system and process failures that would affect the availability of the information.

Be aware that information availability is related to systems but also to processes. When a data integration process fails and the system can't self-correct it, the end result for business users and processes is the same as a database or reporting system outage. The availability analysis should be done from source to consumer.

Manage the Information Lifecycle

In today's environment, with its explosion of information governed by privacy and audit laws, developing processes to manage the information lifecycle through the creation, use, archival, and destruction of data is imperative to control costs and reduce risks.

Consider the management of this lifecycle as you build your information roadmaps and define the flexible information infrastructure architecture. Trying to retrofit systems to deal with such complex areas of information management is time-consuming, drives costs up, and affects the expected business return.

You should also understand that information is more than just data coming out of systems. Processes and users rely on rules to interpret information, so consider

managing and governing your business rules as well. Not only must the rules be consistent across the organization, but they also need to be consistent over time.

Validate Quality and Accuracy

Define the importance of collaboration between business and IT to maintain the accuracy of data requirements and business rules. One way to bring stakeholders together is to focus on data governance. Data governance requires the orchestration of people, process, and technology to enable the organization to leverage trusted information as a strategic asset.

As with the other deliverables from the information agenda, critical success factors include the following:

- An executive sponsor to lead, promote, and communicate the project's benefits
- Constant and comprehensive communication between IT and the business users
- Identifying performance measures for success to evaluate progress

In Conclusion

Information is a corporate asset and a key source of corporate innovation. It is an asset that requires investment to inventory and manage, but it can provide returns hundreds to thousands of times the original investment.

Building an information strategy based on the goals, initiatives, and objectives of your company can be a significant step in developing your roadmap to success through information innovation. Driving an information agenda will enable your organization to deliver trusted, dynamic, enterprise information using an orchestrated combination of people, proven processes, and technology to produce breakthrough results in its ability to identify and manage risk, meet regulatory requirements, speed new products and services to market, and enhance customer service.

Driving Business Optimization with Trusted Information

Leveraging IBM InfoSphere and Cognos Solutions

Business optimization requires information that is accurate, complete, in context, and actionable. Achieving this level of trusted business information requires transforming, reconciling, and maintaining information, and delivering it in real time to the people, processes, and applications that need it. IBM InfoSphere and Cognos lead the industry by offering the breadth of capabilities required for the end-to-end management and delivery of trusted business information to ensure competitive success.

BILL WONG IBM InfoSphere executive responsible for delivering and supporting solutions built on InfoSphere Warehouse and InfoSphere Balanced Warehouse

DR. GUENTER SAUTER IBM senior architect working on trusted information, information integration, master data management, and information as a service

BRIAN BYRNE IBM cross-product architect addressing IBM Information Server, Cognos, MDM Server, and Industry Models

DAN WOLFSON IBM Distinguished Engineer and the chief architect and CTO for the Information Platform and Solutions segment

HARRIET FRYMAN Product marketing manager for the Cognos business intelligence and performance management platform

PAULO PEREIRA Chief Information On Demand architect for IBM Information Management

DR. WILLIAM O'CONNELL IBM Distinguished Engineer and CTO for the Data Warehousing Solutions segment

PHIL DOWNEY Technical Marketing Program Manager for IBM InfoSphere Warehouse

REX WIEDERANDERS Worldwide Technical Sales Enablement Manager for IBM InfoSphere Warehouse

ALAN MEYER Marketing manager for IBM InfoSphere Warehouse for System z

MC Press Online, LP
125 N. Woodland Trail
Lewisville, TX 75077

Price: $14.95 US/$16.95 CN

ISBN 978-1-58347-088-6
51495
9 781583 470886