Beyond Big Data

Using Social MDM to Drive Deep Customer Insight

Martin Oberhofer ▪ Eberhard Hechler ▪ Ivan Milman
Scott Schumacher ▪ Dan Wolfson
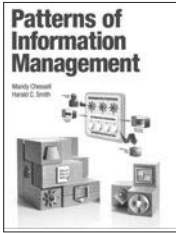
Related Books of Interest

Patterns of Information Management

By Mandy Chessell and Harald C. Smith
ISBN: 978-0-13-315550-1

Use Best Practice Patterns to Understand and Architect Manageable, Efficient Information Supply Chains That Help You Leverage All Your Data and Knowledge

Building on the analogy of a supply chain, Mandy Chessell and Harald Smith explain how information can be transformed, enriched, reconciled, redistributed, and utilized in even the most complex environments. Through a realistic, end-to-end case study, they help you blend overlapping information management, SOA, and BPM technologies that are often viewed as competitive.

Using this book's patterns, you can integrate all levels of your architecture—from holistic, enterprise, system-level views down to low-level design elements. You can fully address key non-functional requirements such as the amount, quality, and pace of incoming data. Above all, you can create an IT landscape that is coherent, interconnected, efficient, effective, and manageable.
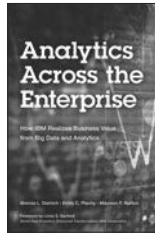
Analytics Across the Enterprise
How IBM Realizes Business Value from Big Data and Analytics

By Brenda L. Dietrich, Emily C. Plachy, Maureen F. Norton
ISBN: 978-0-13-383303-4

How to Transform Your Organization with Analytics: Insider Lessons from IBM's Pioneering Experience

Analytics is not just a technology: It is a better way to do business. Using analytics, you can systematically inform human judgment with data-driven insight. This book demystifies your analytics journey by showing you how IBM has successfully leveraged analytics across the enterprise, worldwide. Three of IBM's pioneering analytics practitioners share invaluable real-world perspectives on what does and doesn't work and how you can start or accelerate your own transformation.

Whatever your industry or role, whether a current or future leader, analytics can make you smarter and more competitive. *Analytics Across the Enterprise* shows how IBM did it—and how you can, too.

Related Books of Interest
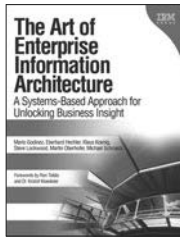
The Art of Enterprise Information Architecture

### A Systems-Based Approach for Unlocking Business Insight

By Mario Godinez, Eberhard Hechler, Klaus Koenig, Steve Lockwood, Martin Oberhofer, and Michael Schroeck
ISBN: 978-0-13-703571-7

Architecture for the Intelligent Enterprise: Powerful New Ways to Maximize the Real-Time Value of Information

Tomorrow's winning "Intelligent Enterprises" will bring together far more diverse sources of data, analyze it in more powerful ways, and deliver immediate insight to decision-makers throughout the organization. Today, however, most companies fail to apply the information they already have, while struggling with the complexity and costs of their existing information environments.

In this book, a team of IBM's leading information management experts guide you on a journey that will take you from where you are today toward becoming an "Intelligent Enterprise."
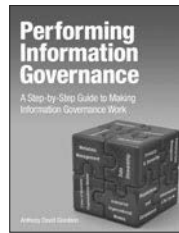
Performing Information Governance

A Step-by-step Guide to Making Information Governance Work

By Anthony David Giordano
ISBN: 978-0-13-338565-6

Make Information Governance Work: Best Practices, Step-by-Step Tasks, and Detailed Deliverables

Using case studies and hands-on activities, Anthony Giordano fully illuminates the "who, what, how, and when" of information governance. He explains how core governance components link with other enterprise information management disciplines, and provides workable "job descriptions" for each project participant.

Giordano helps you successfully integrate key data stewardship processes as you develop large-scale applications and Master Data Management (MDM) environments. Then, once you've deployed an information asset, he shows how to consistently get reliable regulatory and financial information from it.

Performing Information Governance will be indispensable to CIOs and Chief Data Officers…data quality, metadata, and MDM specialists…anyone responsible for making information governance work.

Listen to the author's podcast at:
ibmpressbooks.com/podcasts

Related Books of Interest
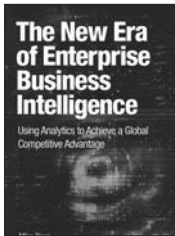
The New Era of Enterprise Business Intelligence:

Using Analytics to Achieve a Global Competitive Advantage

By Mike Biere
ISBN: 978-0-13-707542-3

A Complete Blueprint for Maximizing the Value of Business Intelligence in the Enterprise

The typical enterprise recognizes the immense potential of business intelligence (BI) and its impact upon many facets within the organization—but it's not easy to transform BI's potential into real business value. Top BI expert Mike Biere presents a complete blueprint for creating winning BI strategies and infrastructure and systematically maximizing the value of information throughout the enterprise.

This product-independent guide brings together start-to-finish guidance and practical checklists for every senior IT executive, planner, strategist, implementer, and the actual business users themselves.
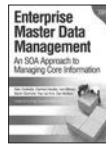
Enterprise Master Data Management

An SOA Approach to Managing Core Information
Dreibelbis, Hechler, Milman, Oberhofer, Van Run, Wolfson
ISBN: 978-0-13-236625-0

Decision Management Systems

A Practical Guide to Using Business Rules and Predictive Analytics
Taylor
ISBN: 978-0-13-288438-9

IBM Cognos Business Intelligence v10

The Complete Guide
Gautam
ISBN: 978-0-13-272472-2

IBM Cognos 10 Report Studio

Practical Examples
Draskovic, Johnson
ISBN: 978-0-13-265675-7
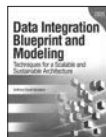
Data Integration Blueprint and Modeling

Techniques for a Scalable and Sustainable Architecture
Giordano
ISBN: 978-0-13-708493-7

This page intentionally left blank

Beyond Big Data

Using Social MDM to Drive Deep Customer Insight

This page intentionally left blank

Beyond
Big Data

Using Social MDM to Drive Deep Customer Insight

Martin Oberhofer, Eberhard Hechler, Ivan Milman, Scott Schumacher, and Dan Wolfson

To my wife Kirsten and my sons Damian and Adrian—
You are the great love and joy of my life.

—Martin Oberhofer

To my wife Irina and my sons Lars and Alex—
Who so greatly encouraged me in this endeavor.

—Eberhard Hechler

To my wife Janie, for making my heart dance with joy.
To Jenica and Ryan, for showing me how a heart can grow.
And to Bill, for joining the club.

—Ivan Milman

I would like to dedicate this to my wife Julie,
who supports me in everything I do.

—Scott Schumacher

To my family, especially my wife Danelle, for their
understanding and support.

—Dan Wolfson

Contents

Foreword I

Every leader makes decisions. And every decision depends on information. That's been true whether someone has led a company, a government, an army, a team, or a household.

The ultimate value of the information technology industry has never been about chips, computers, and software. The industry has always sought to help leaders know with confidence all that has happened, is happening, and might happen to every aspect of the enterprise. But the ante is upped by the volume and variety of information and the velocity of decision making. We're entering the age of Big Data where knowledge and expertise are minimal stakes for survival, and the traditional data centers are becoming the coffins. In this new era of Big Data, continuous reinvention, relevance, and engagement are all that matter. Context is king, and contextual understanding of those you serve becomes absolutely critical to transforming your organization, industry and profession.

As a result, a systematic approach to engagement is now required. You will need to leverage new technology capabilities and business concepts—this is *Social Master Data Management*—to harness data both inside and outside your organizations. Social, mobile, and data together are empowering people with knowledge, enriching them through networks and spawning expectations for real value in return for their information and services, with enterprises they trust. You will want to personalize every meaningful interaction, make transparent those interactions, and continuously earn the right to serve customers. This demands privacy, security, and trust. You need to use mobile and social to increase speed and responsiveness—and meet customers, partners and employees where they are.

Look, data is becoming the world's new natural resource and it is inspiring organizations to take action differently. This book shares best practices in how to think and act differently about your customer data. I encourage everyone to read it!

Inhi Cho Suh
IBM Vice President
Big Data, Integration & Governance

Foreword II

Personal relationships are key to successful business. What if it was possible to provide your customers with the same personal touch they would expect from their favorite local mom-and-pop store when they visit your website, engage in live chat, or speak with technical support? Intuitively, we all know it is a huge competitive advantage because it would improve the quality of every interaction they have with your company.

In the virtual world, how do you provide a personalized experience when there is no physical interaction with people? You can't read their facial expressions as they view one of your products. You can't watch the body language as they are going through the checkout process. The answer is there is still plenty of information to be gleaned from a person's behavior online. You just need the right tools to capture and understand it.

In today's "always-on" society, people provide a wealth of information about their preferences and interests through the websites they visit, the products they rate, the people they follow, and the online communities they engage in. The challenge, of course, is sifting through all this information and making sense of it. Thankfully, advances in analytics and Big Data technologies are making it possible to sift through enormous volumes of information that is known about prospects and customers as well as providing insight into things that can be inferred from their behavior.

This challenge applies to B2C and B2B companies alike as companies of all types and sizes are striving to adopt a "Business to Individual" approach in all their customer interactions. However, it is compounded for B2B companies that must take the extra step of aggregating the interests and preferences for many individuals in a client company in order to infer the preferences and interests of the company at large.

Customers expect a high degree of personalized interactions at every stage of their relationship with a company. As such, quality information about an individual is needed by all the functions in a company; from the marketer working to deliver personalized messages to a technical support representative on the phone with a customer and every process in between.

In this book, you will learn how Master Data Management and Big Data technologies are being combined to arm you with the tools you need to attack this challenge. You will learn how to combine information you know about people with information you can only infer from their

behavior. Further you will learn how to organize it in a way that will enable you to act on it and provide a personalized touch to your customers through all their interactions with your company.

I encourage anyone wrestling with this challenge to take full advantage of the wealth of information in *Beyond Big Data*. It will equip you with the knowledge you need to successfully take on this challenge.

Brian Mackey
Director, Marketing Transformation in IBM's BT/IT organization

Preface

What Is This Book About?

Social Master Data Management (Social MDM) is the new revolution in business data processing that marries customer and product centricity with big data to radically improve customer experiences and product strategy. Traditional master data management (MDM)—the set of processes, practices, and technologies for creating a single view of common core business objects shared across multiple business processes and multiple systems (such as *customer, product, vendor, and location*)—is widely used by enterprises to improve the marketing, operational, and support processes for their customers. However, the focus of traditional MDM is structured data—and today, valuable information about customers and products is locked inside of vast amounts of unstructured, transactional, and social data such as tweets, blogs, Facebook, email, call center transcripts, call data records, and so on. There has been an explosion in technology like Hadoop and BigInsights to extract that information, but often those efforts have limited reach because they are not tied into the existing insight about customers and products contained in MDM systems.

In *Beyond Big Data: Using Social MDM to Drive Deep Customer Insight*, we explain how the union of social, mobile, location, and master data:

- Creates a richer relationship with existing customers
- Improves how you find and target new customers with the right products
- Delivers deeper understanding of how your customers think and feel about your products
- Brings the immediacy of mobile technology to create new ways to engage with customers

Chapter 1, "Introduction to Social MDM," explains the basic concepts of master data and MDM. It describes how disparate data is linked together as cleansed and standardized master data in a master data management system. We show the typical use cases for MDM of customer care and insight (as well as product catalog management), and then introduce the concepts of how social data can extend and enhance MDM into a more powerful system of Social MDM.

Chapter 2, "Use Cases and Requirements for Social MDM," dives into a set of use cases for Social MDM related to improved customer experience, improved target marketing, deeper product understanding, regulatory issues, and the role of location in Social MDM. We then explore requirements and capabilities for the new types of insight and relationships that can be gleaned through integration of master data and social, mobile, and location data. These relationships cover person to person relationships, person to product relationships, person to organization relationships, and others.

Chapter 3, "Capabilities Framework for Social MDM," describes the different data domains that are in scope of a Social MDM environment, and then gives an overview of the core information capabilities needed to deliver a Social MDM solution.

Chapter 4, "Social MDM Reference Architecture," builds on the capabilities described in Chapter 3, and places them as components in a reference architecture. The reference architecture shows the interaction, layers, and relationships between the components.

Chapter 5, "Product Capabilities for Social MDM," links the capabilities described in Chapter 2, and the architectural components described in Chapters 3 and 4, with actual products and technologies that provides those capabilities.

Chapter 6, "Social MDM and Customer Care," looks in detail at how to create a more compelling customer experience using Social MDM. A specific reference architecture for Social MDM and Customer Care illustrates how to create customer centricity through offline, online and real-time capabilities of analyzing social and other enterprise data, linking it to MDM, and then delivering a more tailored experience through a variety of channels.

Chapter 7, "Social MDM and Marketing," shows how the move from traditional broad-based marketing to target marketing is accelerated through Social MDM. This chapter illustrates how to get a deeper understanding of your customers and products to create compelling offers, and how to create more effective (and different types of) marketing campaigns that yield appropriate offers based on Social MDM, identify influencers to expand the market, and use contextual marketing to deliver the right offer at the right time.

Chapter 8, "Mobile MDM," takes Social MDM in a different direction, by showing how MDM can inform and improve mobile applications, and how Social MDM can incorporate mobile data to improve customer experience and grow employee productivity. This chapter looks at the characteristics of mobile data, and modifies the Social MDM architecture to accommodate mobile data and mobile channels.

Chapter 9, "Future Trends in MDM," reveals how the traditional MDM capabilities for entity resolution and matching can be scaled out and enhanced with a Big Data platform. We also look at an emerging technology in the MDM space known as Semantic MDM. Semantic MDM uses new ways of representing the knowledge we have through MDM and social data along with semantic technology to derive new insights and relationships, giving us a better understanding of our customers. Finally, this chapter looks at the privacy and ethical considerations of how we gather, analyze, and use the Social MDM ecosystem, and what are the ethical considerations we must address at every step of Social MDM projects.

Who Should Read This Book

Beyond Big Data: Using Social MDM to Drive Deep Customer Insight has information and insight for a range of practitioners and roles in the enterprise. For business leaders looking to understand how to combine social and master data to create new business opportunities and improve their existing business, this book has excellent material on MDM, Social Data, business value, privacy concerns, and approaches to the new world of Social MDM.

For technical leaders such as Enterprise Architects, Information Architects, and System Architects, this book explores the technologies and use cases in detail, and specifically includes a reference architecture along with domain-specific guidance about best practices to drive a Social MDM project. It also includes a product mapping that gives direction on which technologies and products to adopt to solve particular scenarios.

What You Will Learn

This book provides a wide-ranging exploration of the business, technical, and ethical landscape of Social MDM. We cover the basic concepts of master data and master data management, and the same concepts for social data. You'll learn how Social MDM mediates the relationship of customers to the business, yielding greater insight about customers (so you can serve them better) and providing better service and value to customers (so they will have a greater incentive to buy more of your products and services).

We look at the different types of insight (cultural awareness, sentiment, detailed customer segmentation, influence of individuals) you can derive from social, underused enterprise, and mobile data and show how that is incorporated into a Social MDM platform. You will understand the architecture and capabilities of a Social MDM system, with a mapping to specific technologies and products. This book articulates how that architecture and those capabilities can be used to drive enhanced customer care and to build advanced marketing campaigns leveraging deep and broad insight of your customers, targeting them with the right offers and incentives (and avoiding the wrong ones). You will learn the new technologies brought about through mobile systems and how that extends and modifies the capabilities of a Social MDM system.

You will get a peek into new technologies to scale out and extend traditional MDM services in entity resolution and linking, as well as semantic technologies that add a learning and reasoning layer on top of Social MDM. Finally, you will be challenged to understand that just because you have all of this data and insight does not mean you have the right to use it. Privacy laws and customer expectations will be at the heart of a socially responsible MDM.

How to Read This Book

Beyond Big Data: Using Social MDM to Drive Deep Customer Insight is logically structured into three sections:

Concepts, Business Value Capabilities, and Ethics: These are targeted at business leaders who want to understand Social MDM and how it differs from traditional master data and

analytics, delve into the new business opportunities derived from Social MDM, explore the capabilities required for Social MDM, and finally, reflect on the legal, ethical, and business implications of using Social MDM.

This section includes the following chapters:

- Chapter 1, "Introduction to Social MDM"
- Chapter 2, "Use Cases and Requirements for Social MDM"
- Chapter 9, "Future Trends in MDM"

Capability, Architecture, and Product Mapping: These chapters are aimed at technical leaders who need to understand the overall technical approach to Social MDM and the roles of the underlying components in the Social MDM architecture. This section includes:

- Chapter 3, "Capability Framework for Social MDM"
- Chapter 4, "Social MDM Reference Architecture"
- Chapter 5, "Product Capabilities for Social MDM"

Social MDM Domains: These chapters are for both business and technical leaders who want to understand the specific technical details about how Social MDM enhances business opportunities in these areas:

- Chapter 6, "Social MDM and Customer Care"
- Chapter 7, "Social MDM and Marketing"
- Chapter 8, "Mobile MDM"
- Chapter 9, "Future Trends in MDM"

Conventions

Following is a short list of key conventions that are used throughout this book:

- **Abbreviations**—Abbreviations are used across the book chapters, where all abbreviations are spelled out when they are used the first time in the book.
- **References**—This book includes quite a number of references for further study, where all references are listed at the end of each individual chapter. This way, you will find relevant information for further study in the context of the topics of each chapter. Footnotes are used to link the relevant statement in the chapter to the corresponding reference(s).
- **Footnotes**—Additional footnotes provide further background information, for example in regards to products or tools mentioned.
- **Italic type**—Key terms, new concepts, and important aspects within a statement, a list, and also in tables are emphasized through use of italic type.
- **Figures and tables**—Figures and tables are numbered consecutively in each chapter.

Acknowledgments

About the Authors

Martin Oberhofer works as Executive Architect in the area of Enterprise Information Architecture with large clients world-wide. He helps customers to define their Enterprise Information Strategy and Architecture solving information-intense business problems. His areas of expertise include master data management based on an SOA, data warehousing, Big Data solutions, information integration, and database technologies. Martin delivers Enterprise Information Architecture and Solution workshops to large customers and major system integrators and provides expert advice in a lab advocate role for Information Management to large IBM clients. He started his career at IBM in the IBM Silicon Valley Labs in the United States at the beginning of 2002 as a software engineer and is currently based in the IBM Research and Development Lab in Germany. Martin co-authored the books *Enterprise Master Data Management: An SOA Approach to Managing Core Information* (IBM Press, 2008) and *The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight* (IBM Press, 2010) as well as numerous research articles and developerWorks articles. As inventor, he contributed to more than 70 patent applications for IBM and received the IBM Master Inventor title. Martin is certified by The Open Group as a Distinguished Architect and holds a master's degree in mathematics from the University of Constance/ Germany.

Eberhard Hechler is an Executive Architect who works out of the IBM Boeblingen R&D Lab in Germany. He is currently on a three-year assignment to IBM Singapore, working as the Lead Architect in the Communications Sector of IBM's Software Group. Prior to moving to Asia, he was a member of IBM's Information Management "Integration and Solutions Engineering" development organization. After a two-and-a-half year international assignment to the IBM Kingston Development Lab in New York, he has worked in software development, performance optimization and benchmarking, IT/solution architecture and design, and technical consultancy. In 1992, he began to work with DB2 for MVS, focusing on testing and performance measurements. Since 1999, he has concentrated on Information Management and DB2 on distributed platforms. His main expertise includes the areas of relational database management systems, data warehouse and BI solutions, IT architectures and industry solutions, information integration, and Master Data Management (MDM). He has worked worldwide with communication service providers and IBM clients from other industries. Eberhard Hechler is a member of the IBM Academy of Technology, the IBM InfoSphere Architecture Board, and the IBM Asset Architecture Board. He coauthored the books *Enterprise Master Data Management* (IBM Press, 2008) and *The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight* (IBM Press, 2010). He holds a master's degree (Diplom-Mathematiker) in Pure Mathematics and a bachelor's degree (Diplom-Ingenieur (FH)) in Electrical Engineering (Telecommunications).

Ivan Milman is a Senior Technical Staff Member at IBM working as a security and governance architect for IBM's Master Data Management (MDM) and InfoSphere product groups. Ivan co-authored the leading book on MDM: *Enterprise Master Data Management: SOA Approach to Managing Core Information* (IBM Press, 2008). Over the course of his career, Ivan has worked on a variety of distributed systems and security technology, including OS/2® Networking, DCE, IBM Global Sign-On, and Tivoli® Access Manager. Ivan has also represented IBM to standards bodies, including The Open Group and IETF. Prior to his current position, Ivan was the lead

architect for the IBM Tivoli Access Manager family of security products. Ivan is a member of the IBM Academy of Technology and the IBM Data Governance Council. Ivan is a Certified Information Systems Security Professional and a Master Inventor at IBM, and has been granted 14 U.S. patents. Ivan's current focus is the integration of InfoSphere technology, including reference data management, data quality and security tools, and information governance processes.

Scott Schumacher, Ph.D., is an IBM Distinguished Engineer, the InfoSphere MDM Chief Scientist, and a technology expert specializing in statistical matching algorithms for healthcare, enterprise, and public sector solutions. For more than 20 years, Dr. Schumacher has been heavily involved in research, development, testing, and implementation of complex data analysis solutions, including work commissioned by the Department of Defense. As chief scientist, Scott is responsible for the InfoSphere MDM product architecture. He is also responsible for the research and development of the InfoSphere Initiate matching algorithms, and holds multiple patents in the entity resolution area. Scott has a Bachelor of Science degree in Mathematics from the University of California, Davis, and received his Master of Arts and Doctorate degrees in Mathematics from the University of California, Los Angeles (UCLA). He is currently a member of the Institute for Mathematical Statistics, the American Statistical Association, and IEEE.

Dan Wolfson is an IBM Distinguished Engineer and the chief architect/CTO for the Info-Sphere 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 and Quality for Big Data including Information Quality Tools, Information Integration, Master Data Management, and Metadata Management. Dan is also CTO for Cloud and Mobile within Information Management, working closely with peers throughout IBM.

Dan has more than 30 years of experience in research and commercial distributed computing, covering a broad range of topics including transaction and object-oriented systems, software fault tolerance, messaging, information integration, business integration, metadata management, and database systems. He has written numerous papers, blogs, and is the coauthor of *Enterprise Master Data Management: An SOA Approach to Managing Core Business Information* (IBM Press, 2008). Dan is a member of the IBM Academy of Technology Leadership Team and an IBM Master Inventor. In 2010, Dan was also recognized by the Association of Computing Machinery (ACM) as an ACM Distinguished Engineer.

Social MDM Reference Architecture

Reference architectures encapsulate architectural best practices harvested and harnessed from a series of implementations. In this chapter, we introduce the Social MDM Reference Architecture regarding its key capabilities based on the capability framework. The primary purpose of this chapter is to enable you to understand the relevant components, their relationships, and interactions for building MDM solutions—specifically for Social MDM use cases.

Introduction

In this chapter, we introduce the Social MDM Reference Architecture (Social MDM RA). We do this by embedding a master data management (MDM) system in the broader enterprise context of operational and analytical systems. With the rise of Big Data in information management, the architectural ecosystem for MDM systems is changing alongside the analytical systems. Social persona information, customer sentiment, etc. is analytics-derived using Big Data analytics. Thus it is not surprising that Social MDM solutions depend on new enterprise information architectures also able to deliver Big Data solutions. We will show you in the Architecture Overview how the ecosystem in which Social MDM solutions live evolved over the last few years by showing you the old and the new environment.

Using a component model comprised of a component relationship diagram providing a static perspective of the key components, as well as some component interaction diagrams providing dynamic views, we introduce the Social MDM RA on a more detailed level.

Architecture Overview

In this section, we first take a look at MDM in the application landscape for an enterprise followed by the introduction of an architecture overview. When we introduce the architecture overview, we first take a look at the ecosystem prior to the rise of Big Data. In a second step, we show how the architecture overview evolved due to the impact of Big Data.

MDM as Central Nervous System for Enterprise Data

Although many MDM implementations historically focused on operational use cases, with the rise of Social MDM, an MDM system truly becomes the central nervous system for the enterprise, as shown in Figure 4.1. It is connected to the operational landscape as well as to a broad range of analytical applications. A key observation in Figure 4.1 is that in many cases the connections are bidirectional because with Social MDM, the MDM system becomes a core essential part of the operational fabric. For example, although social media analytics might enrich a particular customer record with insights gleaned from unstructured sources such as social media, customer interaction logs from the call center, and so on, the starting point for that analysis is the customer records that define a "search scope" to the analysis. Similarly, with self-service capabilities to update their master data record exposed to the customer through various operational channels, the link between operational applications and MDM becomes more and more bidirectional where a couple of years ago many MDM systems were fed with a consolidation style architecture pattern.
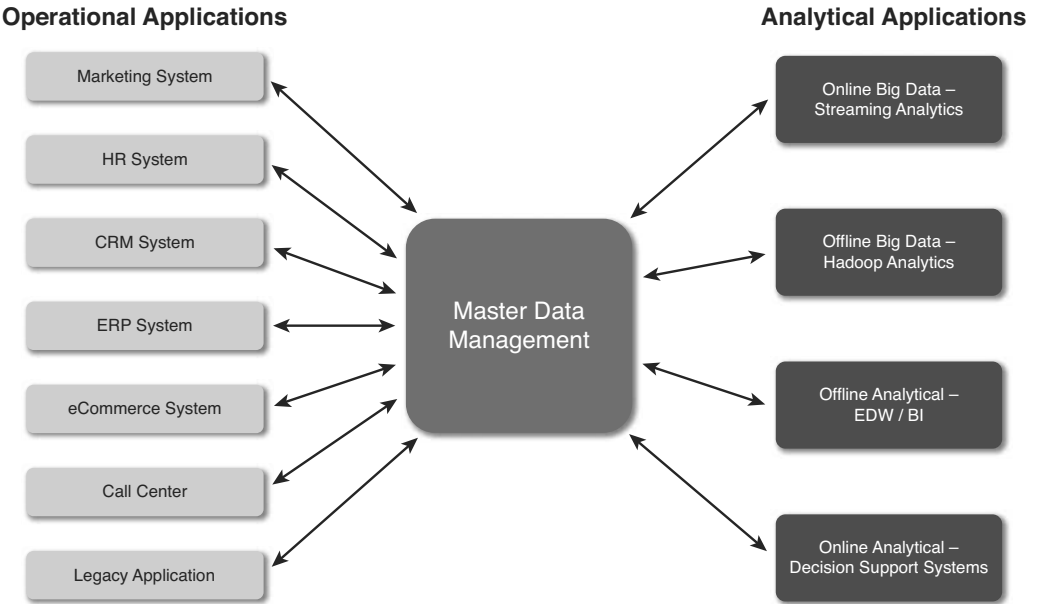
Figure 4.1   MDM—the central nervous system for enterprise data

MDM: Architecture Overview

Now that you have a better understanding of the functional scope of the discussed capabilities in the previous chapter, let's switch gears to implementation architecture. A few quick words regarding nomenclature will help to more easily convey key messages in the drawings. A *functional area* is a collection of related subsystems delivering a major IT function. A *technical capability* is a specialized type of technology performing a specific role; we introduced those relevant to us in Chapter 3. With information provisioning as an example, there are collections. In this example, it is a collection of mechanisms for locating, transforming, or aggregating information from all types of sources and repositories. A *zone* is a scope of concern describing a usage intent for a particular cross-cutting service. It has associated requirements and governance that any system in the zone must adhere to. Figure 4.2 shows iconic examples we use for these concepts in the drawings.

Figure 4.2 Nomenclature

To understand what is changing with Social MDM, we first need to understand common deployment architectures today, such as shown in Figure 4.3.

In Figure 4.3, you can see two types of capabilities:

- **Technical capabilities introduced in Chapter 3:** Examples include (but are not limited to) Master Data Hubs, Reference Data Hubs, and so on, which are technical capabilities introduced in the Information Engine capability layer in the category Managed Operational Data Hub. Other capabilities are grouped in functional areas; for example, the Analytic Sources Area is composed of the capabilities in the Data Server category from the Information Engine capability layer as well as some analytical functions from the Insight capability layer.

- **Technical capabilities external to the capabilities defined in Chapter 3:** These are primarily well-known IT systems such as customer relationship management (CRM) applications.

In the functional area of traditional sources on the left side in Figure 4.3 are the sources for master-data-comprised third-party data sources such as Dun & Bradstreet, as well as operational applications such as customer relationship management (CRM), enterprise resource planning (ERP), human resources (HR), supply chain management (SCM), supplier relationship management (SRM), and eCommerce. In a typical enterprise, some of these applications are packaged from vendors like SAP and Oracle, or from software as a service (SaaS) providers like Workday and Salesforce.com, or custom-built applications.
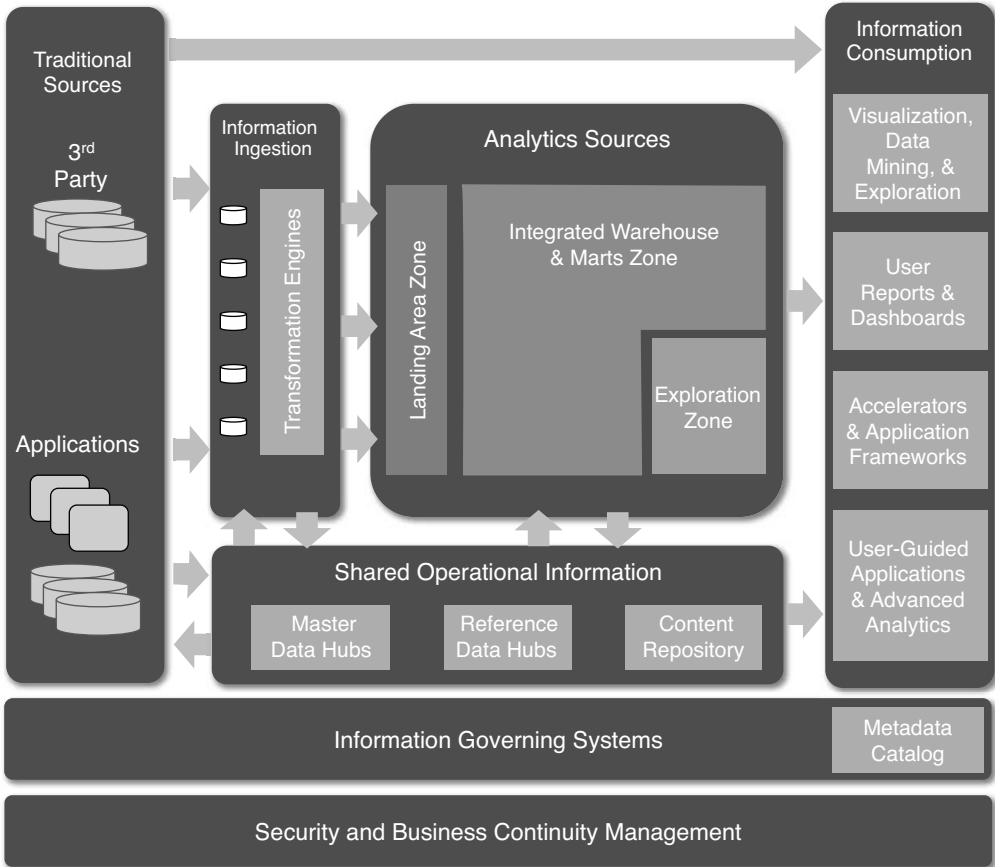
Figure 4.3   Architecture overview—a traditional viewpoint

The functional area of information ingestion has transformation engines providing, for example, ETL or CDC capabilities. Using these transformation engines master data can be moved from the sources to MDM or from MDM into the data harmonization processes feeding the analytical sources. The MDM system resides in the functional area of shared operational information systems alongside Reference Data Management and Content Management Systems. The name "Master Data Hubs" is intentionally plural for two reasons: first, commercial software vendors historically provided Master Data Management software for a single domain only, such as for a customer or product, creating the two disciplines customer data integration (CDI) and product information management (PIM). Early adopters of MDM sometimes implemented multiple MDM products for different purposes, from the same or different vendors resulting in multiple master data hubs. Today, many MDM software vendors provide multi-domain MDM software often reducing the number of distinct hubs. Multiple Master Data Hubs can also be the result of a

merger and acquisition where both companies have an MDM system already. Yet another reason could be that the company adopted different MDM software solutions from different vendors to address different MDM requirements. The functional area for analytical sources is composed of the landing zone where the data harmonization for operational data stores and data warehouses located in the integrated warehouse and marts zone is done. For exploratory analytics such as pattern detection, a dedicated exploration zone exists. For the functional area of information consumption where business users consume information, the figure shows various well-known technical capabilities such as data mining and reporting. For governing the information architecture, the functional area of information governing systems provides a metadata catalog storing business technical and operational metadata, among other capabilities. The functional area of security and business continuity management provides necessary security features for controlling and auditing information access as well as features for backup and restore, high availability, disaster recovery, maintenance, and so on.

With the rise of Big Data, the implementation landscape changes to reflect the new sources and capabilities available as shown in Figure 4.4.
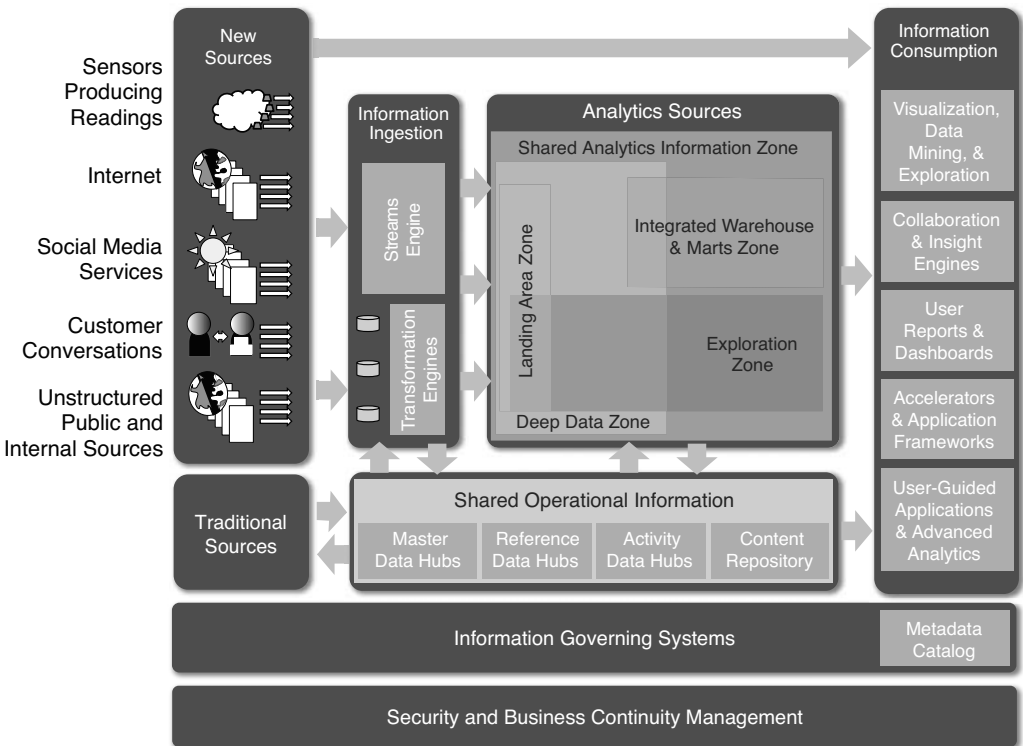
Figure 4.4 Architecture overview—impact of Big Data

Major changes in key functional areas are:

- **Data sources:** A whole new group of data sources has emerged. As internet-connected sensors and devices become more common (often called the Internet of Things), more information can be collected, integrated, and analyzed to improve operational efficiencies and quality of life across a number of areas. Examples include instrumentation for food transport ("farm to fork"), utility networks (smart water/gas/electricity networks and smart meters), and smarter homes as just three examples implementing sensors producing data at an unprecedented rate and massive volume. New kinds of unstructured content sources have also emerged including blogs and wikis. Social media sources grow at a rapid pace as well, and examples include Facebook, Twitter, LinkedIn, and Yelp.

- **Information ingestion:** A new technique known as streams processing has emerged to address new use cases where data is produced at speeds and volumes too large to actually persist all the data. A streams engine can apply real-time analytics as information is created to make timely decisions and to selectively store the most interesting information.

- **Analytical sources:** A new zone of deep analytics is added—the location of new analytical capabilities based on the MapReduce paradigm, as we will see. With a Hadoop platform to implement a Map-Reduce platform allows you to land the data, perform possibly some cleansing, do some analytics, and persist the results of the analytics which might be also moved to a DW. With such a system, you would have all historic and current data. This possibly changes the DW procedures because instead of archiving of the DW you can simply delete because the full history is still in the Hadoop platform.[1] The second major change is that the consumption of information is radically simplified, creating a true shared analytics information zone.

- **Information consumption:** New techniques of collaboration and new insight engines appear as novel technical capabilities. Examples include new matching engines to search for duplicates and nonobvious relationships, pattern mining, and natural language analytics.

- **Information governing systems:** Major functional enhancements include the extension of the metadata catalog to enable a broad class of users to find and provision the information they need from across the variety of systems and zones.

Figure 4.5 shows the architecture overview from an Social MDM perspective. Integration and analysis of new sources of information, especially social media sources, is one of the most striking changes from Figure 4.5. Another key change is the introduction of activity hubs.

---

1. More details on this topic can be found in [1].
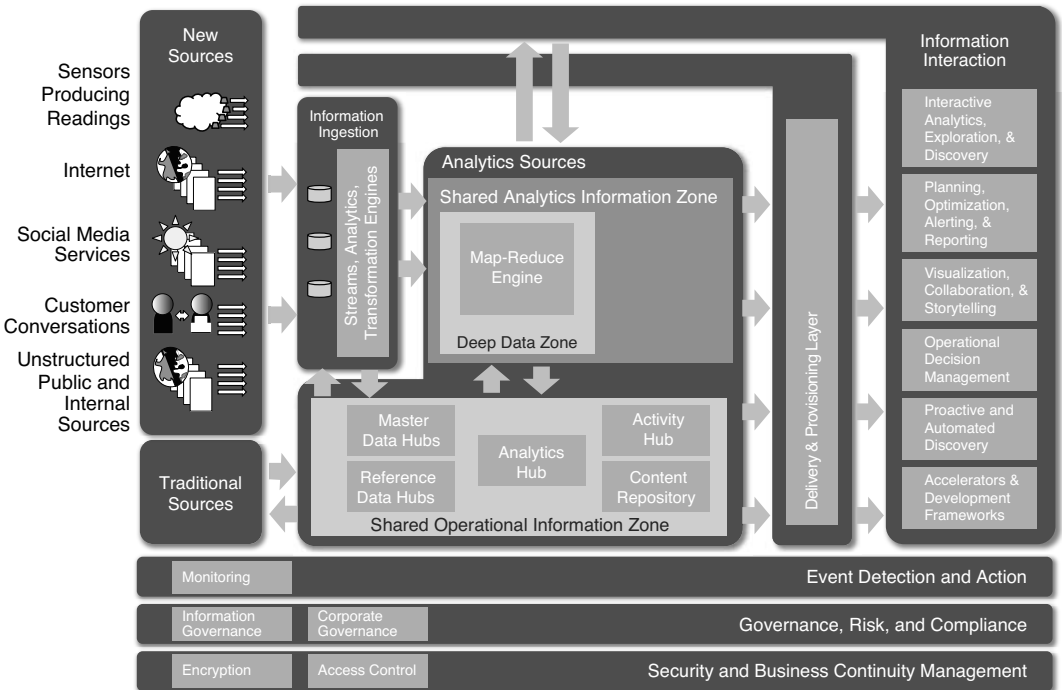
Figure 4.5 Architecture overview—focus on Social MDM

Component Model

The architecture overview in the previous section is useful to provide enterprise architects, IT managers, and business users a conceptual perspective of key capability areas for a Social MDM architecture, however, it is not detailed enough to understand how the components in this architecture interact with each other. For this, a component model is necessary.

In the following sections, we explore the Social MDM RA on a functional component level using a component relationship diagram. A component relationship diagram provides a static view of the relationships between the components. We then walk through a sample component interaction diagram to demonstrate the dynamic interaction of the components. Component interaction diagrams will be used in subsequent chapters to describe various angles of the Social MDM RA from different architectural perspectives.

Component Relationship Diagram from an Enterprise SOA Perspective

You may have read [1], which provides an enterprise Master Data Management reference architecture aligned with an SOA-based enterprise architecture blueprint.[2] Figure 4.6 shows an updated version of the component relationship diagram published in [1].
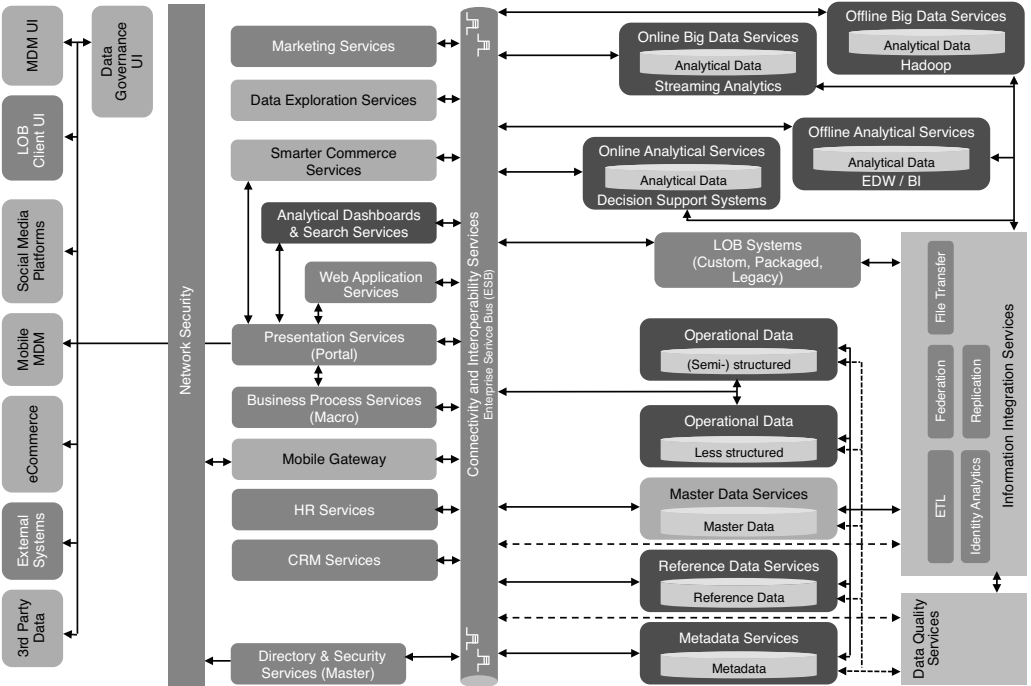
Figure 4.6   Social MDM component relationship diagram from an SOA perspective

The major differences are as follows:

- **Simplification:** Since the previous component relationship diagram was published as part of the MDM RA, roughly six years have passed. Working with many clients during this time, we found that many components have become commoditized as SOA has matured. Following are some examples:

  - The DMZ zone with a reverse proxy pattern, firewalls, and so on has been collapsed into a single network security component.

---

2. You probably noticed that several authors of this book coauthored the previous book, so we decided to provide you with this perspective as well as an evolution to the previous version.

- Different types of LOB systems have been summarized into a single component.
- The subcomponents of the MDM component are not explicitly shown anymore. They are well understood and common knowledge in the MDM practitioner community at this point.
- Various messaging gateway components and the interconnectivity and interoperability component have been merged into a single ESB component.

- **Changes in data domains:** At the time the previous MDM RA was created, reference data was subsumed as part of master data. With the lessons learned since then, we see it now as a separate domain with domain-specific functionality to be appropriately managed, thereby justifying a component in its own right.

- **New third-party data sources:** Over the years, we have worked with enterprises that expressed some degree of dissatisfaction with commercial third-party vendors. Examples of issues include lack of completeness, cost, and and to some extent staleness of the data. Social media sources are perceived to address some of these aspects better but at the same time with the risk of some uncertainty and fuzziness. Also, additional processing is required to make these sources usable from an MDM standpoint. For these reasons we moved this into a different component.

- **New use cases:** The expansion of banking services to support mobile devices is one of several drivers for Mobile MDM—a capability that also requires new components, which we added to the component model.

- **The impact of Big Data for analytics:** As indicated with the data domain introduction, Big Data use cases broaden the scope of analytical capabilities, which we represent with a functional processing capability set with dedicated components.

Component Relationship Diagram for Social MDM from an Information Architecture Perspective

Although the SOA-centric component relationship diagram of the MDM RA remains a very useful reference architecture, for Social MDM, we felt the urgent need for a more information-centric view. Figure 4.7 shows the Social MDM RA component relationship model from an information-centric perspective. As you will notice, most of the important components from the MDM RA remain but some components have been removed or combined to allow us to focus on Social MDM. For example, network components have been removed and others have been combined such as portal, Web, and eCommerce.
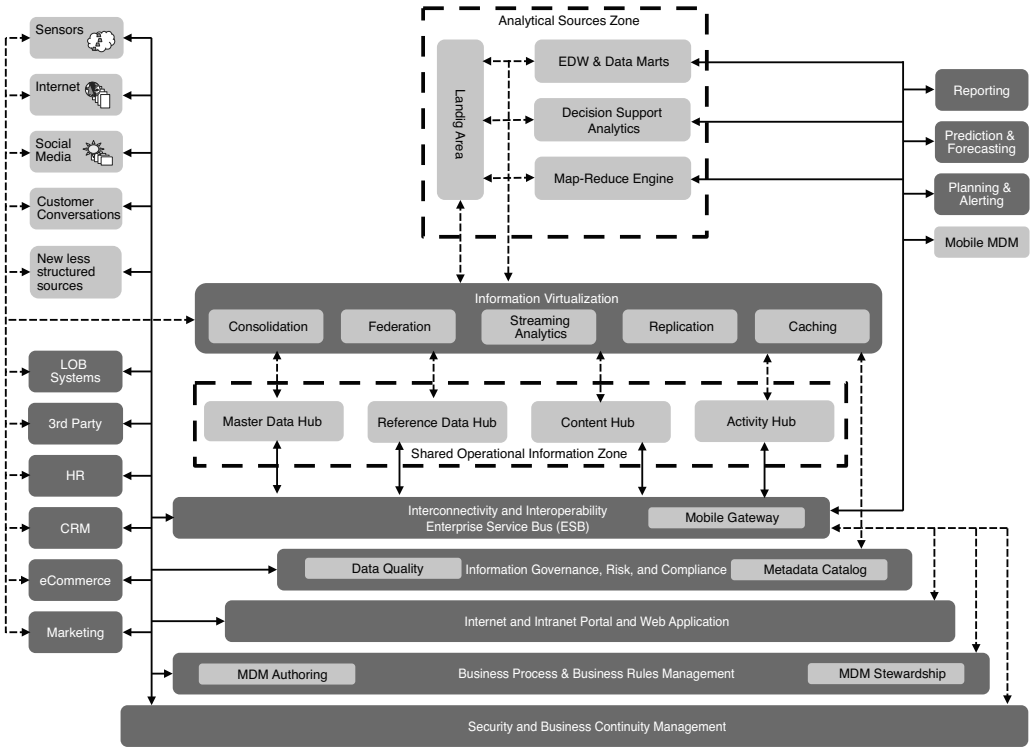
Figure 4.7   Social MDM component relationship diagram from an information architecture perspective

On the left side of the figure, you can see the key operational data sources—traditional ones like CRM, HR, and marketing systems as well as new sources like social media platforms. The Master Data Hub and the Reference Data Hub are enterprise-wide (and in some cases, cross-enterprise) shared information assets. Through the information virtualization layer and the ESB, they are connected to all relevant IT systems to provide seamless access to these critical information assets. The information virtualization layer provides technical capabilities for information access, consolidation, replication, and so on. In industries such as banking, the digital transformation brings new banking services on mobile platforms like Android or iOS for smart phones. The ESB has been extended to include a mobile gateway that exposes services to mobile platforms. A mobile gateway provides the ability to map a Web service interface to a REST service interface to make the service consumable for Android devices. Additionally, the mobile gateway provides Mobile MDM capabilities independent of the target device environment—thereby supporting native Android, iOS, or even HTML5-based applications. Within the analytical sources zone, the

landing area enables data harmonization prior to analytical use in other components such as the DW and data mart. A new component is the MapReduce engine, which we will explore in depth in Chapter 5. The information in the analytical sources is consumed as analytical data through technical capabilities such as reporting, prediction and forecasting, and enterprise planning.

For a Social MDM solution to work, additional common infrastructure components are necessary. The information governance, risk, and compliance component, for example, provides capabilities for data quality (for example, address standardization services) and a metadata catalog (for example, to manage business terms for master data entities alongside logical and physical data models of the Master Data Hub). The Internet and Intranet portal and web application component delivers functionality to surface master data such as employee or intranet employee dictionary applications or customer master data self-service functionality for being able to notify an organization about address or contact detail changes. The business process management component delivers business process functionality for authoring and stewardship processes of master data. Typical master data process examples include new product introduction, hierarchy maintenance, account creation, or duplicate suspect processing. Like other IT solutions, Social MDM depends on appropriate security and business continuity infrastructure. An MDM solution usually has very demanding requirements regarding availability (after all, master data is required by many of the mission-critical operational and analytical applications) as well as security (for example, enterprises need to protect their customer master data or face reputational damage as well as possible legal consequences).

Component Interaction Diagram

Whether or not a component relationship diagram represents a good component model can be assessed by a use-case-driven validation through component interaction diagrams. In subsequent chapters, you will see a lot more component interaction diagrams for various use cases. In this section, however, we start with a basic use case in a business-to-consumer (B2C) scenario where we would like to add influencer scores to the customer master data records in the Master Data Hub. To determine influencer ratings, we need to analyze information coming from sources such as videos on YouTube, blogs, wikis, forums, posts on Facebook and Twitter, and so on. Analyzing these sources requires a broad range of analytics such as voice-to-text or video analytics on the raw data, followed by deeper analytics on topic and sentiment detection. These analytics may be implemented as a sequence of Map-Reduce jobs based on a social analytics library. The output of these jobs is written to a well-modeled social media mart where predictive analytics are deployed to determine influencer scores. In addition, the data derived from social media can also be used for the matching of social personas to customer records in the Master Data Hub. Figure 4.8 shows the end-to-end component interaction diagram for this use case.
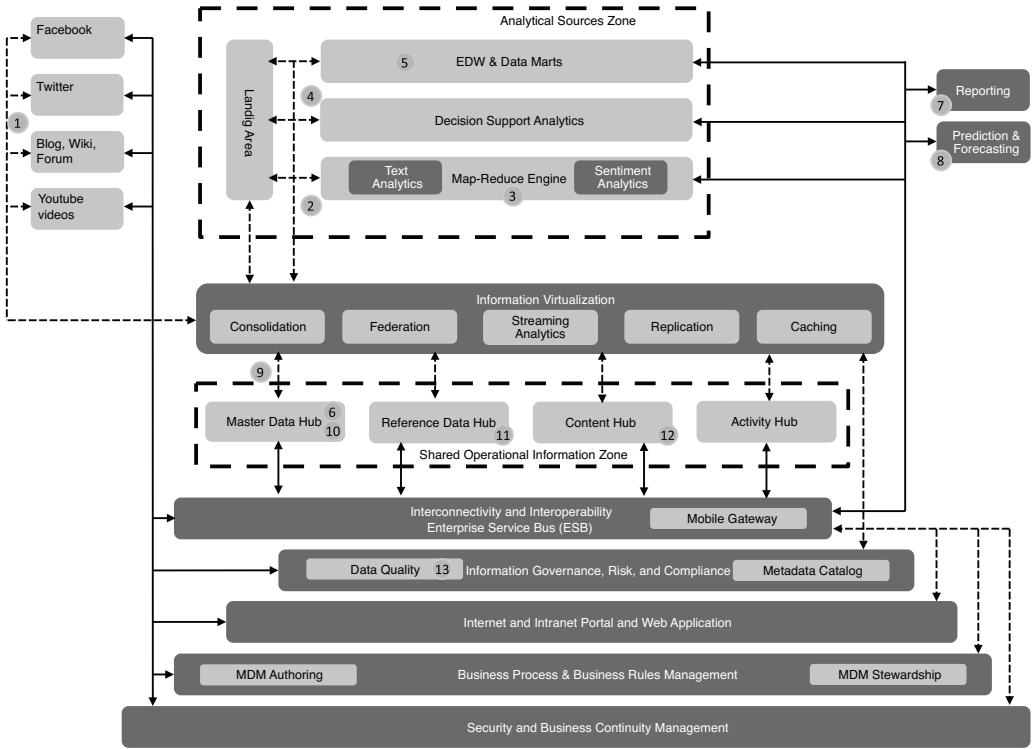
Figure 4.8   Component interaction diagram for enriching master data with influencer scores

Step 1. In this step, connectors consume data feeds from the new information sources such as Facebook and Twitter. Depending on software selection, the connectors might be part of the information virtualization component or the MapReduce engine. For monitoring ongoing marketing campaigns, events, or 1:1 customer engagement opportunities, information is accumulated frequently—usually every few minutes—for some situations, for example where we are integrating customer service with social media, we may receive updates continuously.

Step 2. In this step, the raw data is loaded into the MapReduce engine.

Step 3. Configured as a series of analytical operations, the MapReduce engine executes a broad range of analytics. Initially, "document-centric" analytics strive to identify concepts, social personas, authoring location, demographics, behavioral patterns, and sentiment. Subsequently, topic detection analytics are applied, including correlation and assignment of individual mentions of concepts to topics. Leading MapReduce engines can execute this analytical sequence in a matter of minutes.

Steps 4 and 5. After the MapReduce engine finishes, the results are moved to a social media mart. Depending on the database used for the social media mart as well as the software for the MapReduce engine, this could be as simple as a flat file produced by the MapReduce engine that gets loaded through a bulk load interface into the database. Of course, alternatively, the flat file might also be loaded with ETL software (the consolidation engine in the information virtualization component) that may also be used to simultaneously restructure and enhance the information as it is loaded.

Step 6. Using the matching capability of the Master Data Hub,[3] we can now determine which social persona might correspond to a customer record. There might be social personas with no match in the Master Data Hub or vice versa where a customer uses different social personas on different social platforms.

Step 7. The outcome of the matching task can be visualized with reports using the reporting component.

Step 8. Using predictive analytics with appropriate models and scoring functions, we can compute influencer scores.

Steps 9 and 10. The influencer scores can be moved from the social media data mart to the Master Data Hub using capabilities such as ETL (part of the consolidation functionality) from the information virtualization layer. As a result, the master data records in the Master Data Hub are enriched with the influencer scores from the social media platforms.

In addition to these core steps of the use case, there are additional options for consideration. Influencer scores might change over time from a more coarse-grained to a more fine-grained level or vice versa. Influencer categories based on lower and upper thresholds might be defined through reference values. If the influencer score analytics suggest a change in these categories, the corresponding reference value sets might require an update in the Reference Data Hub (step 11). Possible social media sources might contain additional opportunities to enrich master data beyond the influencer scores. There might be additional contact, address, or other demographic information found—that with cleansing through appropriate data quality services (13) could be added to the matched master data records as well. For example, if a match is found for a social persona, there might be pictures or documents available from the social media that could be persisted in a Content Hub (step 12) and linked as unstructured master data to the master data record in the Master Data Hub.

---

3. IBM ported the MDM matching engine to the Hadoop Map-Reduce engine. This version of the IBM MDM matching engine is known as BigMatch. If such technology is available, then the comparison of social personas using a matching engine can happen within the Map-Reduce environment. If a match is found with a feasible degree of similarity, then depending on the certainty of the match, either a link is established or the customer profile gets enriched.

Subject-Oriented Integration

With an understanding how the Social MDM Reference Architecture works, we want to now investigate how we can make master data available from a consumer standpoint. Provisioning the subset of information based on consumers' social context requires the capability to perform *subject-oriented integration* for Social MDM solutions. Let's look at some examples:

Ubiquitous Internet availability anywhere anytime on mobile devices allows people to interact with corporate IT in novel ways. The salesperson meeting a customer's contact person wants to refresh his memory on that person's current social context by doing a lookup from a mobile device in the car just a few minutes before entering a meeting at the client's site. A member of the support organization looking at a product defect reported by someone from the customer's site would like to know if that person is possibly already deeply frustrated. In this case, the support engineer would like to see if that person who opened the product defect report already posted negative statements online about the product or posted questions in forums related to the problem reported. Having that context available might affect the support engineer's style of communication with the person who opened the product defect report and may reduce the time of analyzing the problem (there is no need to explore causes and possible solutions that have been identified as not helpful by the person opening the product defect report in forum discussions). These are just two examples illustrating how different consumers in different social contexts have different perspectives on the same 360-degree complete social master data based on their role.

Before we can integrate information, we need to *understand* it first, and this also applies to subject-oriented integration. Although information integrations have been built for a long time, in many cases this has been done in an *ungoverned* way. For example, an extract-transform-load (ETL) developer might have looked at source and target models and just built an ETL program based on a mapping of source to target attributes. The semantics of each field of the source and target data models, the relationship of these models to certain business entities and functions, and user roles consuming the data on the target are not captured in many cases. For a while, an inhibiting factor was the lack of appropriate metadata software to manage business, technical, and operational metadata and data lineage; and impact analysis functionality was not part of enterprise information integration platforms. This changed in the past several years with commercial metadata management solutions now available, fully supporting the creation of enterprise glossaries where technical (for example, logical and physical data models, mappings, and data profiling rules) and operational metadata can be attached to business metadata (for example, terms and policies). With this metadata functionality available through an enterprise glossary, it's now possible to seamlessly define a *subject* composed of:

- A term describing the subject from a business perspective
- An assigned owner (for example, an information steward) who is responsible and accountable for the data asset described in the term

- Policies governing how the information asset related to the term has to be managed in terms of data quality, security, retention, privacy, and so on, and implementation rules linked to these policies used for enforcement

- Technical metadata linked to the term expressing how and where the data asset is stored, how it is related to other assets (for example, mappings), and permissible value ranges (for example, through linkage of applicable reference data tables)

- Operational metadata such as results of enforced security constraints, measured data quality, and so on

With the subject defined, it's now possible to provide a catalog for information assets in the enterprise. Various users in the enterprise looking for information assets can now use this catalog of information assets. However, on a high level, we see two very distinct use cases. First, based on requests from the business, technical users can develop the integration to deliver the necessary subject considering all constraints attached to it for the consuming users from a consumption point of view. Second, business users can browse the catalog of information assets by subject, and if state-of-the-art *self-service capabilities* provided through the enterprise information integration platform are available and enabled, they can generate the integration for the subject of interest to the desired consumption point without having IT personnel involved.

More details on this topic of subject oriented integration can be found in [2].

Conclusion

In this chapter, we provided an introduction to the Social MDM RA. We provided an architecture overview and a component model perspective, organizing the functional scope of the Social MDM RA into architectural deliverables. We demonstrated how the Social MDM RA works by using a component interaction diagram showing how it works for a specific use case. We then explained how to use the concept of subject-oriented integration to provision master data for Social MDM solutions from a consumption viewpoint.

In the next chapter, we explore software solutions which can be used to deliver the discussed components in the component model in more depth.

References

[1] Dreibelbis, A., E. Hechler, I. Milman, M. Oberhofer, P. van Run, and D. Wolfson. *Enterprise Master Data Management: An SOA Approach to Managing Core Information* (IBM Press, 2008).

[2] Chessell, M., Smith, H.: *Patterns of Information Management*. (IBM Press, 2013).

This page intentionally left blank

Index