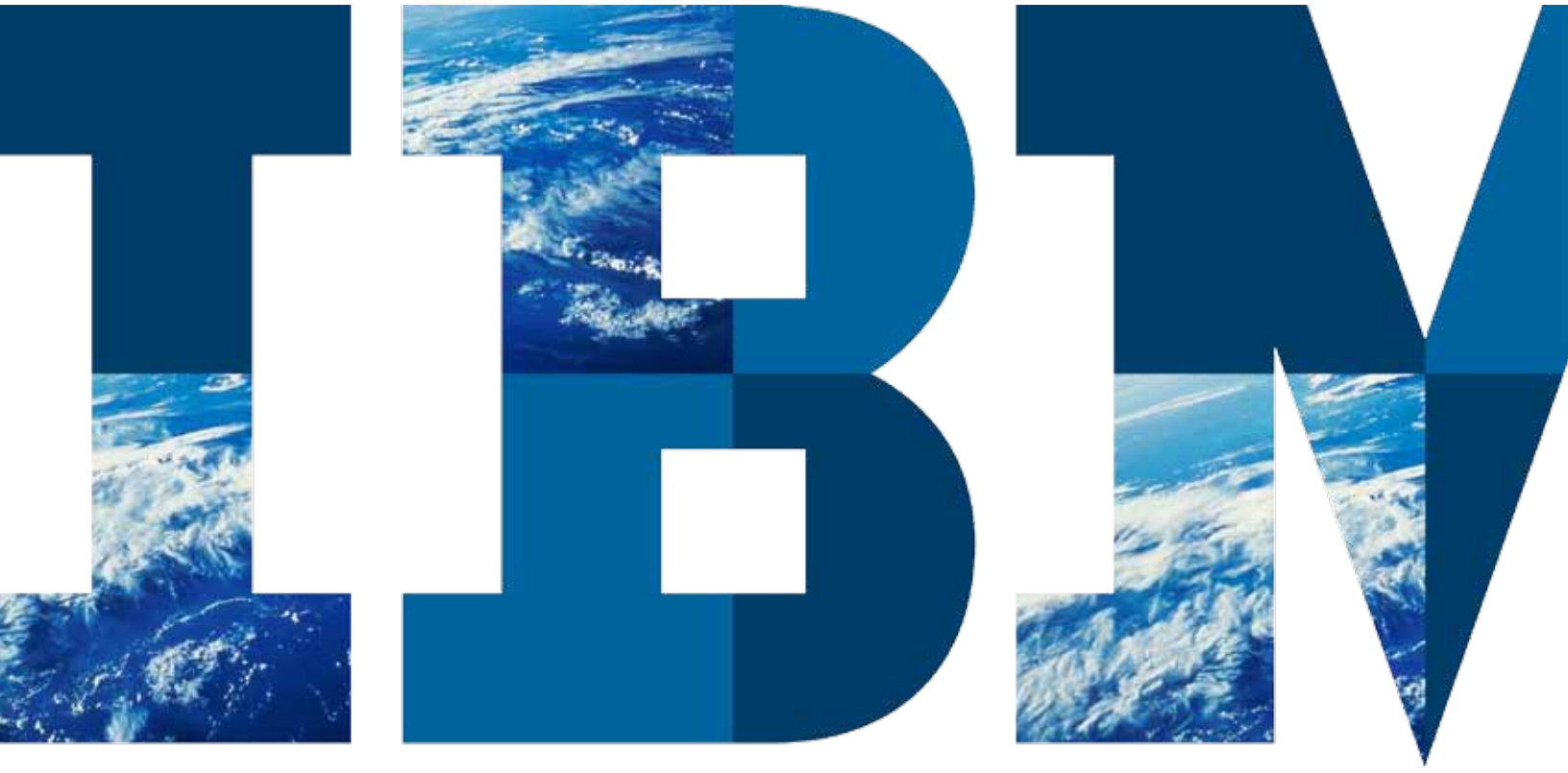
IBM Software
White Paper

Master data management: The key to leveraging big data



IBM

Many regard big data as the next major transformative force in enterprise IT. However, its four defining characteristics can make it difficult to tap its full potential:

1. **Volume:** According to the Digital Universe study, in 2009, the world had stored 800,000 petabytes of information. By 2020, that will rise to about 35 zettabytes.[1] A zettabyte is one trillion gigabytes.
2. **Variety:** Big data is often unstructured—for example, text comments, images or video.
3. **Velocity:** Big data generated by sensors and instruments can stream in at rates 1,000 times greater than IT departments typically handle.[2]
4. **Veracity:** Big data can be unreliable and tough to verify. For example, you might know which customers posted angry comments about your restaurant on Facebook, but their claims might be incorrect or exaggerated.

As a result of these four characteristics, organizations need specialized tools and methods to handle and leverage big data. Master data—the most trusted subset of all data—can help enterprises make sense of the enormous volumes of big data at their disposal. Master data is the basis for a shared repository of information that is used across an enterprise; it pertains to any entity in a business, from customer records to products, and has been culled and verified to become a consistent "golden record" about that entity. Master data management (MDM) encompasses the tools and discipline of governing and leveraging master data.

By combining big data with verified and trusted master data, companies can gain new insight on customers or products, or derive immediately actionable information (see Figure 1).

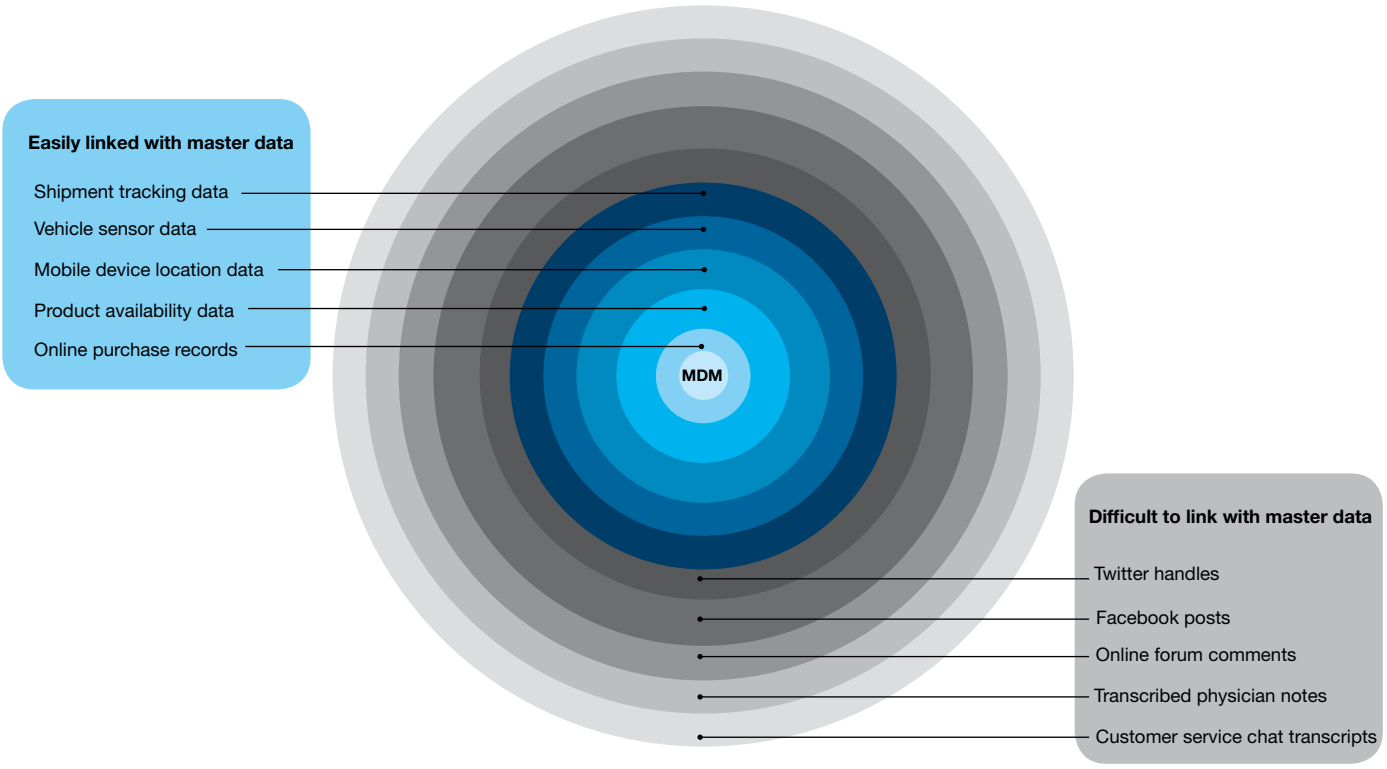


Figure 1: MDM helps turn big data into trusted data through governance and verification.

Governing big data: Making the connection with MDM

Big data is governable, but on a continuum. Few organizations, if any, will ever process and analyze all the big data they accumulate; it will likely continue to grow and diversify ahead of attempts to map and analyze it. However, by identifying, classifying and using as much of the data as possible, organizations can gain valuable insights that would otherwise go undiscovered.

CIOs struggling to make effective use of big data are likely to have an MDM problem, because information about your customers buried within big data needs to be linked to or related to existing customer information already captured in traditional structured applications. Without linking it back to your master data, new customer or product information extracted from big data ends up being yet another silo of data.

Increase veracity of big data by linking it to master data

MDM's well-defined approach augments the value of big data by giving it structure and meaning. MDM tools can match products with traits derived from web comments and social network postings. MDM also gives context to machine-derived data—such as the data provided by mobile devices about the times and places consumers buy specific products.

MDM helps improve the 360-degree customer view by enhancing master data with unstructured content. For example, Facebook postings may tell a hotel chain that a high proportion of its business guests have too many children for standard reward rooms. The hotel can then respond by providing reward privileges on larger suites for these high-value customers, potentially increasing their loyalty.

MDM can also deepen understanding of customers with structured data that is captured from a high-velocity data stream. Consider the case of insurance companies that evaluate driving habits—good and bad—with information from vehicle sensors, and use that information to adjust insurance rates accordingly. In addition, MDM helps organizations uncover relationship links that enhance existing knowledge about customers and open up opportunities to gain new business or save costs. For example, by establishing that two separate customers are part of the same household, the company can avoid sending duplicate mailings to the same address.

Big data is a resource that can help organizations identify, understand and take action to address large-scale problems such as customer-centric service delivery, fraud, healthcare efficacy and one-to-one, real-time promotional offers to mobile users. And by accurately joining big data to golden records in an MDM system, organizations increase the level of trust in the information they use to make critical business decisions.

"The coming of the big-data era is a chance for everyone in the technology world to decide [whether they come to work for a paycheck or to change the world]…this era will bring the biggest opportunity for companies and individuals in technology since the dawn of the Internet."

—Rob Thomas, vice president, IBM Software Group[3]

Focus on adding value to unstructured data

Without an enterprise-scale, strategic approach to leveraging big data, data growth can become a runaway train. MDM helps control the momentum of incoming data, and tools that apply MDM to big data help expand the universe of verifiable facts about customers, vendors and products.

Every big data strategy should focus on enabling a company to handle customers more profitably. For example, MDM can help a hotel chain link TripAdvisor reviews on its hotels (unstructured data) to specific customer profiles (structured data). This hotel chain can then ascertain with reasonable certainty that some of its top-tier customers have sizable Twitter, FourSquare or Facebook followings. It can then single out those customers for a thank-you phone call with some carefully chosen perks.

The IBM InfoSphere MDM portfolio

IBM® InfoSphere® Master Data Management (InfoSphere MDM) software is a complete, powerful and operational MDM solution, including both transaction-oriented MDM and collaborative authoring and workflow capabilities to handle all domains, implementation styles and use cases across various industries. An MDM-powered application toolkit delivers business value by creating governance applications through pre-built blueprints and widgets that you can embed within existing applications. Pre-built data models are extensible for any domain and optimized for MDM for rapid time-to-value—or organizations can import existing data models or build data models from scratch. In addition, InfoSphere MDM is designed to support large transaction loads and high-availability operational environments.

Best practices for increasing the trustworthiness of unstructured data with MDM

Set up processes to make large data sets more valuable—for example, motivate customers to match their unstructured data to master data. Software tools may make a fuzzy match between Twitter IDs and customer records, but the customer can give you a 100-percent-positive match much more quickly.

MDM can be a major element of data governance. A realistic approach to governing big data involves:

- **Matching what can reliably be matched to master data**
- **Classifying unstructured data as accurately as possible**
- **Using common metadata between big data and MDM initiatives**
- **Deciding what levels of quality and veracity are good enough for your organization, and then focusing on the big data that meets those levels**
- **Applying the same master data privacy and security policies to the big data that you've correlated to master data**

Combining the power of InfoSphere MDM and big data capabilities

IBM has a variety of solutions focused on analyzing big data to help organizations uncover previously unavailable insights and use them to support and inform decisions across the business (see Figure 2). Combining the power of InfoSphere MDM with these IBM big data solutions creates a valuable connection: big data technology can supply insights to MDM, and MDM can supply master data definitions to big data.

Analyzing big data at rest

By analyzing large volumes of stored data, organizations can discover patterns and insights that allow them to optimize processes and profitability. For example, one healthcare insurance provider analyzed more than 400 million claims, looking for potentially deadly drug interactions. Text from patient phone calls is intricately nested and often mentions medications obtained elsewhere or without a claim form. For example, the standard drug name might be Diflucan, but it could be called Flucanol, Batacan or yeast medicine in a transcript. The company's internal system took more than 100 hours to perform this analysis.

With IBM InfoSphere BigInsights™ matching transcripts to master data records of medications, however, analysis time dropped to just 10 hours for more than 6 TB of data.[4] InfoSphere BigInsights also excels at extracting structured data from text analysis and matching it to master data. These capabilities pave the way for a new class of applications for customer experience analysis, social media analysis and fraud detection.

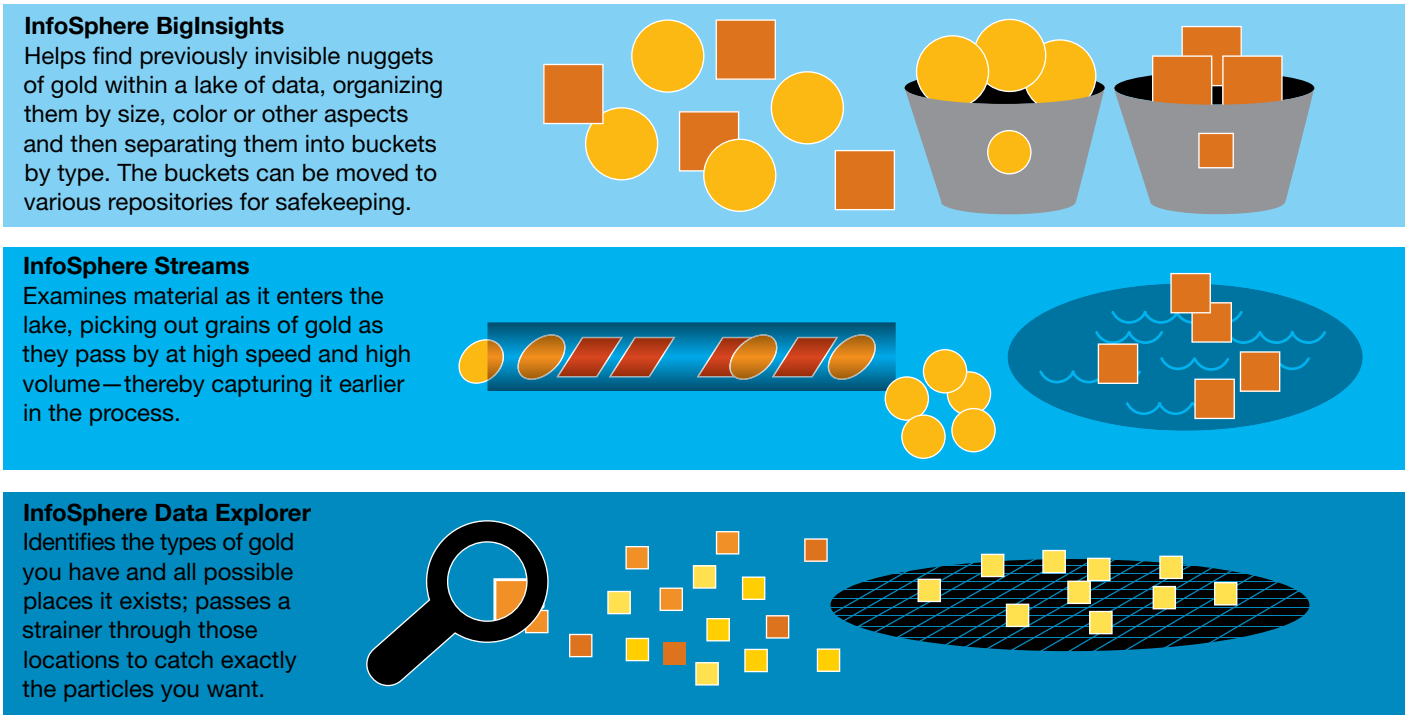


Figure 2: As in gold mining, you often have to sift through a lot of worthless material to find the valuable nuggets—but the right tools make the process more efficient. The IBM platform for big data offers several methods of finding "gold" insights in a lake of big data, depending on business need.

Analyzing big data in motion

With certain kinds of data, there is no time to store it before acting on it. This data is often (but not always) generated by a sensor or other instrument. Speed matters for use cases such as fraud detection, emergent health care and public safety, where gaining insight into data in real time—and determining whether to act according to predefined rules—can be critical.

High-velocity, high-volume data calls for in-motion analytics. Streaming data may arrive linked to a master data identifier (a phone number, for example). In a neonatal ICU, sensor data is clearly matched to a particular infant. Financial transactions are unambiguously matched to a credit card or Social Security number. However, not every piece of data that streams in is valuable—which is why it should be analyzed by a tool such as IBM InfoSphere Streams before being joined to master data.

InfoSphere Streams analyzes large data volumes with micro-latency. Rather than accumulating and storing data first, Streams analyzes data as it flows in and identifies conditions that trigger alerts (such as outlier transactions that a bank flags as potentially fraudulent during the credit card authorization process). When this occurs, the data is passed out of the stream and matched with master data for better business outcomes. InfoSphere Streams generates a summary of the insights derived from the stream analysis and matches it with trusted information, such as customer records, to augment the master data.

Analyzing mobile data in motion

Master data adds strategic value to big data generated by mobile communications. Globe Telecom, a leading telecommunications company in the Philippines, uses InfoSphere Streams for real-time analytical processing (RTAP) of data about call, recharge and churn activity. The company used this information to identify better promotion patterns for individual customers. Delivering a recharge offer at the optimal moment has had a profound impact on its business.

Linking mobile big data and MDM also opens up a range of cross-selling and up-selling possibilities based on personal preferences that can be linked to master data and prioritized, based on time-day and geospatial location.

Catching the customer: The importance of content for promotional offers

At 6 p.m. on a Saturday, John Sanchez parks his SUV near the beach and uses pay-by-phone parking for his meter. Based on information gleaned from Facebook postings and its own master data, the telecom provider believes John is a 40-year-old father. It asks John, via text, if he would like a discount code for a family-sized pizza from a restaurant just steps from his parking spot. If John's profile indicated he is 21 years old and single, however, the coupon might be for a nearby club or tattoo parlor.

Big data from mobile devices often enables organizations to make definitive connections to a specific individual. Logons via tablet from coffee shops and airports generate staggering amounts of geolocation data—with corresponding potential gains for retailers that can make effective use of this data.

Explore big data to uncover patterns and new insights

IBM InfoSphere Data Explorer complements InfoSphere BigInsights and InfoSphere Streams in using master data to leverage big data. This platform for navigating and filtering large amounts of data in nearly any format offers a visual dashboard that joins master data about an individual with that person's detailed transaction history from emails, social media, packaged applications and more (see Figure 3). This data can help sales staff, particularly inbound and outbound telemarketers, connect with prospects on a more individual level and achieve better sales rates.

Designed to handle both social and sensor-generated data, InfoSphere Data Explorer lets users specify an "axis" and then searches for matching data that it can fuse onto existing master data. With master data as the starting point, InfoSphere Data Explorer can also place real-time activity streams into context for analysis.

Intelligence, law enforcement and fraud analysts need to leverage all relevant pieces of information to connect the dots and make well-informed decisions instantly. InfoSphere Data Explorer works from known master data to give intelligence agencies a unified view of all their information sources to deliver visibility, transparency and insight.

Figure 3: InfoSphere Data Explorer pulls in master data information to populate certain areas, presenting it along with information from other sources to create a complete view of an entity, such as a customer.

Big data and MDM: Better together

Any enterprise that accumulates large volumes of fast-moving structured and unstructured data needs a well-defined approach to store, analyze and leverage it, as well as gauge its trustworthiness.

IBM InfoSphere MDM and big data tools are keys that unlock the value of big data. They help companies analyze big data, match it to master data and derive actionable information to inform decisions and extract insights. InfoSphere MDM complements big-data analytics tools and helps organizations deliver trusted information to win sales, satisfy and retain customers, improve operations and increase compliance.

For more information

To learn more about how MDM helps companies utilize big data and the IBM InfoSphere MDM platform, please visit:

- **ibm.com**/software/data/master-data-management
- **ibm.com**/bigdata

© Copyright IBM Corporation 2012

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
November 2012

IBM, the IBM logo, ibm.com, BigInsights and InfoSphere are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at **ibm.com**/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

[1] IDC Digital Universe Study, sponsored by EMC, May 2010.

[2] Hopkins, Brian. "Big data: How enterprise architects can capitalise on future opportunities." http://www.computerweekly.com/feature/Big-data-How-enterprise-architects-can-capitalise-on-future-opportunities

[3,4] "Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data," McGraw-Hill, 2012. **ibm.com**/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&appname=SWGE_IM_DD_USEN&htmlfid=IML14297USEN&attachment=IML14297USEN.PDF

Please Recycle