Synergizing Master Data Management and Big Data

The strategic value of master data management (MDM) has been well documented. Yet for companies with mature MDM systems, the complexities of big data can be overwhelming — challenging them to distill and make meaning from the information that flows into and through their organization.

Executive Summary

Today’s data-intensive environments are populated with multiple databases and database domains that continue to proliferate within organizations — making it difficult to manage the silos of data that house information related to customers, products, assets, suppliers, financial matters and more.

Master Data Management, or MDM, is fundamental to cleansing, standardizing, matching, merging and governing these assets, and assuring that they remain uncorrupted, up-to-date, and afford a “single version of the truth” across domains. At the same time, the practices associated with MDM can be stymied by the ever-growing amounts of big data – petabytes and even exabytes – that originate from different sources and include both semi-structured and unstructured formats.

While big data is top of mind for IT executives, it presents very different challenges compared with master data management. Understanding and comparing the characteristics of MDM and big data and determining how each informs the other can result in fresher, deeper and more actionable insights to support the right strategies, enhance decision making and uncover new competitive advantages.

This white paper offers a way for IT organizations to achieve greater synergy between MDM and big data, and arm leadership with the insights and foresights they need to make more informed business decisions.

Two Approaches, Different Capabilities

Master data management allows an organization to link all of its critical data to one file, called a master file, which provides a common point of reference. When properly executed, MDM can streamline data sharing among personnel and departments, and facilitate computing in multiple-system architectures, platforms and applications. From this viewpoint, MDM is both a process and a technology.

MDM is internally focused on the data at hand, and works hard to provide a single version of the truth for all master data – producing reports, providing dashboards and analytics to gauge its progress, and automating the process of...
improving data quality. MDM works best with relatively low volumes of data, perhaps involving customers, products and geographies. It “likes” structured data that is highly organized and readily searchable by straightforward search-engine algorithms (social media need not apply).

Big data operates in a different, vastly larger universe. It accumulates information relentlessly — constantly discovering new behaviors and insights.

One of the challenges today’s businesses face is preserving the investments they’ve made in MDM systems that unify and optimize discreet silos of information while managing to cope with the ever-increasing influx of big data. From the process-oriented perspective of MDM, this can seem overwhelming. The question is: Can big data actually support the MDM process rather than undermine it?

The answer lies in finding where and how big data complements MDM, and where and how MDM complements big data. It is not an either/or situation. A synergistic blend of the two processes and technologies can help ensure that the benefits of both are retained and enhanced within the enterprise. Going forward, organiza-
Going forward, organizations can benefit enormously from a combination of structured MDM domains that are informed by the predictive, analytical capabilities of big data.

Key Integration Concepts
There are key areas where big data can inform and enhance the MDM process. For the purposes of this paper, we will focus on data storage, data integration, data search/analysis and data security.

Data Storage
MDM employs a master data hub for various data domains, be they databases of customers, products, customer locations, geographies or other types of data slices. Big data comprises large volumes of both structured and unstructured data flowing from blogs, social media, clicks, live event streaming, machine data, telemetry and more. Linking algorithms and APIs from the MDM system to a big data platform can allow data from the two to reside in one place where they can be mutually supportive.

Data Integration
Even though MDM processes maintain a “golden” copy of a master data record, that record can be continuously informed by relevant data from the big data platform. The MDM hub should have the ability to integrate with big data — using identifiers and reference keys to retain what’s useful and discard what’s not. Meanwhile, big data distributions should utilize speedy, massive parallel processing — allowing for the discovery, matching and linking of master data records.

Here, big data has the ability to supplement and enrich MDM’s golden master record. Semi-structured and unstructured data can be parsed in a variety of ways using sentiment, semantic and virtualization technologies. This kind of “smart” parsing is improving all the time — bringing with it powerful analytic capabilities that merge and load appropriate new information into the MDM master file — enriching it greatly.

Semi-structured and unstructured data can be parsed in a variety of ways using sentiment, semantic and virtualization technologies.

Data Search/Analysis
Being able to find the appropriate data when required is key. MDM uses its own brand of indexes within the master data hub to enable faster searches. A big data hub can be linked with the master data hub to support real-time search capabilities. Again, MDM APIs can combine search capabilities across both hubs to consolidate and integrate results. At the same time, big data platforms’ APIs perform their own searches, and link functions of the master data hub as part of high-volume streaming. The right tools can search for and link all data types, whether structured, semi-structured or unstructured.

Data Security
Cybersecurity is a central concern for today’s enterprises. Even though master data is encrypted and retained according to policy guidelines, information originating from big data must also provide accurate, personally identifiable information, since other confidential sets of attributes are encrypted and masked. The integration of MDM and big data platforms is thus essential — covering data security at rest, in motion, and in use.

Pursuing and Gaining Maturity
Maturity refers to those points at which both MDM and big data can most effectively find synergy. Achieving it requires a central level of understanding from both standpoints to make the two systems fully cooperate and become mutually supportive (see Figure 2, next page). With this in mind, it may be best to start small and grow from there via the following stages:

1. Conduct pilot programs of both systems to explore synergy in bits and pieces. This could involve some lines of business or business units that recognize the value of how big data can inform their “golden” and highly valued data hub managed by an MDM platform.

2. Catalog and manage data sets across structured, semi-structured and unstructured data. In this stage, MDM and big data integration focuses on silos across business units, with standardized definitions and usage pertaining to diverse data types.

3. Optimize both MDM and big data. This involves facilitating the onboarding, availability and consumption of all data types between the two platforms and across all business units. At this point, diverse data is mature enough to deliver insights throughout the organization.
4. **Achieve MDM/big data integration through the analysis and predictability of all data domains.** At this stage, new concepts and insights derived from the master data set are discovered by analyzing their synergies with big data. Now, all data — from both the MDM platform and the inflow of big data — is available to answer tactical and strategic questions pertaining to all data types.

**Practical Use Cases**

Consider the practical benefits of achieving synergy between the “golden” data set maintained by a master data platform and the ever-flowing, ever-growing amounts of big data:

**Social Media**

Today, organizations have the ability to continually gain insight into their customer service, their market and their products from comments on social media. This information can typically be found in the structured data contained in completed Web forms. Companies are also discovering and analyzing social posts to gain more insight into customer preferences, their brand sentiment, gaps in customer service, and existing customers’ likes and dislikes. With full maturity and synergies, all of this information can link to the master data hub — extending and enriching the data already on hand.

**Segmentation**

The modern world of marketing began with targeting customers by broad characteristics such as gender and geography. Today, demographics are considered a fairly rudimentary way of segmenting prospective customers. Real-time Web activity, clicks, video views, devices/sensors and social media, among many other “digital fingerprints” (or Code Halos, as we call them), fill the gaps that exist in the central MDM data domain. Beyond historical data analysis, organizations can now use advanced analytics to foresee customer expectations and needs.

**Security**

As stated earlier, cybersecurity is now a primary focus of database management. Risk management — not absolute prevention of any and all cyber-attacks but rather, mitigation of risk down to acceptable levels — is the new byword in cybersecurity and management. Here, big data analysis is central, with real-time streaming of transactions, logs and sentiment data that allows customers to detect any irregularities that might signal potential fraud. The real-time component is key — enabling companies with cybersecurity risk management plans to react immediately to minimize exposure.

---

**Steps to Achieving Greater Synergy**

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pilot and Prove</td>
<td>Synergies exist in pilots with some proven concepts, and with an exploration of use cases.</td>
<td>Synergistic value can be explored in bits and pieces, since only some lines of business or business units are willing to experiment or can benefit from a piece-meal approach.</td>
</tr>
<tr>
<td>2. Manage</td>
<td>Data sets are cataloged and managed across structured, semi-structured and unstructured data, with governance of people, processes and tools.</td>
<td>The adoption of big data and MDM integration breaks down silos throughout business units, with well-conformed and standardized data definitions and usage across diverse data types.</td>
</tr>
<tr>
<td>3. Optimize</td>
<td>Onboarding, availability and consumption of all data types between big data and MDM across business units is available, with little delay in answering business queries</td>
<td>Governance of diverse data is mature enough to deliver data and insights across business units in a more efficient manner, with proper cataloging and lineage.</td>
</tr>
<tr>
<td>4. Analyze and Predict</td>
<td>More time is spent on discovering newer concepts and insights from the synergies achieved between master data and big data sets.</td>
<td>Organizations have all the data they need to answer tactical and strategic questions across diverse data types.</td>
</tr>
</tbody>
</table>

Figure 2
Internet of Things
The Internet of Things (IoT) is reaching maturity in a variety of applications. Any device connected to the Internet can be instrumented to feed information back to manufacturers about behaviors, usage, breakdowns, security issues, system failures and more. The IoT is here to stay, and only through real-time big data streaming, which continuously informs an MDM central data domain, can industry take advantage of it.

Database Synergies
Just because an MDM-managed hub is considered “golden” doesn’t mean there isn’t more to discover within it. Information flowing from big data affords the opportunity to further analyze and examine what is in the master data domain to determine relationships between master data, enterprise data and external data. This can lead to unexpected findings within the central MDM hub, including the discovery of unique identities and relationships.

The 360-Degree View
The MDM hub is just the starting point. It defines master data, but then analyzes big data sources, often unstructured, through social media, customer transactions, etc., to discover new relationships, hierarchies, intent and brand sentiment, for example.

Looking Forward: Recommendations for Success
Cognizant combines a passion for client satisfaction with technological innovation and deep industry and business-process expertise. We are committed to enabling your company to make better use of big data. But for this to happen, big data must “understand” the master data view and its different domains.

To successfully achieve synergy between master data management and big data, we advise organizations to:

- **Determine business value and feasibility.** Assigning value to an initiative (cost savings, new opportunities, upselling/cross-selling and risk mitigation) is a key consideration when identifying possible use cases that can deliver immediate wins. Don’t boil the ocean. This is not a theoretical exercise; rather, it recognizes how immediate ROI can be achieved from these synergies.

- **Start small.** Develop a single pilot program that explores one narrow concept. Using all necessary resources, identify a line of business or business unit that needs to enhance its basic “golden” database with big data. Then, take a staged approach to increase the chances of success from this pilot.

- **Develop business drivers.** Remember, a primary goal is to advance the prospects of your business. Therefore, enhance your central MDM domain with rationalized sets of big data to support a particular objective and set targets in areas such as customer service, product development, new markets, etc.

- **Be ready to reassess.** While striving to increase business value, sometimes it’s best to stop, reassess and start again. Discover, learn and succeed.

- **Focus on the right things.** While internal efficiencies are always important, stay focused on what all of this implies for your customers’ own success, which is your success. We firmly believe that combining master data management with big data can go a long way toward accomplishing both.
About the Author

Ajay Raina is a Principal Architect (Director) in Cognizant’s Enterprise Information Management business unit, where he oversees engagements in the banking, financial services, healthcare and life sciences segments, among other key industries. His forte is establishing stability, optimization and modernization of enterprise information architectures for data management and analytics initiatives. In accomplishing this, Ajay provides strategic oversight, thought leadership, delivery guidance, technology enablement and solution definitions – blended with leading and proven practices in information management initiatives. Ajay has 20-plus years of information management experience in leading data warehousing, MDM, big data and analytics engagements. He can be reached at Ajay.Raina@cognizant.com.

About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process outsourcing services, dedicated to helping the world’s leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 100 development and delivery centers worldwide and approximately 217,700 employees as of March 31, 2015, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world. Visit us online at www.cognizant.com or follow us on Twitter: Cognizant.