IBM InfoSphere Big Match for Hadoop

Accurately connect structured and unstructured customer data for deep insights and effective decision-making

Highlights

- Analyze and match diverse structured and unstructured customer data sets to enhance the utility of analytic output

- Achieve high levels of matching accuracy across large data volumes for improved decision-making

- Realize fast time to value using tried-and-tested matching techniques

The era of big data is upon us. Organizations are eager to leverage massive volumes of information and new types of data to generate unprecedented insights and improve business outcomes. Customer data is at the heart of many of these initiatives. In fact, nearly 50 percent of all organizations with big data projects have identified customer-centricity as a top objective.[1] The impact is clear for organizations that foster customer-centricity using big data analytics:

- Personalized service experiences powered by a more complete understanding of customer preferences
- Increased effectiveness of sales and marketing efforts through comparisons of customers and large target market behaviors
- More attractive offerings developed using product or brand sentiment analysis

Unsurprisingly, IT leadership is facing mounting pressure from business stakeholders to deliver better customer insights and business outcomes. However, the limitations of existing infrastructure often stand in the way of success. Whether it's overwhelming data volumes, the inability to effectively manage new types of customer data, or inaccuracies in matching internal and external customer data sources, the barriers to big data customer analytics can appear insurmountable.

IBM® InfoSphere® Big Match for Hadoop helps organizations analyze massive volumes of structured and unstructured customer data to gain deeper, more comprehensive customer insights. The scalable platform can enable fast, efficient linking of data from multiple sources to provide complete and accurate customer information—without the risks of moving data from source to source. For greater flexibility, InfoSphere Big Match for Hadoop supports a wide variety of platforms, including the latest versions of IBM Open Platform with Apache Hadoop and Hortonworks Data Platform.

Achieve a high level of matching accuracy

Accurately matching customer records has a dramatic impact on analysis. Correctly identifying that multiple fragmented or duplicate records represent the same customer can improve service levels, boost marketing campaign effectiveness and accelerate sales execution. However, given the volumes and variety of big data, matching is no small task.

InfoSphere Big Match for Hadoop plays a critical role in enhancing data accuracy by streamlining the process of sorting out incorrect, duplicate and related data. The solution leverages proven, pre-configured algorithms to compare data elements and assign similarity scores, or weights, which take into account the probability of certain data elements matching based on the distribution of data in the actual data set. Scores are assigned based on the likelihood of a match, and then links are established between records. To strengthen customer-centric big data

initiatives, InfoSphere Big Match for Hadoop also matches and links customer profile data with social media data.

All matching is performed natively in Hadoop, which moves processing closer to the data and minimizes data movement for an order-of-magnitude performance improvement over traditional relational database system processing.

With the ability to automate accurate matching at a fraction of the time and resources of other matching engines, InfoSphere Big Match for Hadoop meets strict business requirements for precision and efficiency.

Support big data types and formats

The diversity of analytically relevant customer data is expanding rapidly, creating limitless opportunities for competitive advantage: in some cases, organizations are systematically gleaning product usage information from support forums to identify product deficiencies. In others, analytic teams are empowering customer service departments with geolocation data to personalize customer interactions. And in yet more situations, companies are capturing sentiment information from social media feeds to guide many strategic product and service-related decisions.

By addressing the variety of data types and formats in these examples—structured and unstructured, governed and ungoverned—InfoSphere Big Match for Hadoop offers an efficient approach to unearthing unique insight (Figure 1).
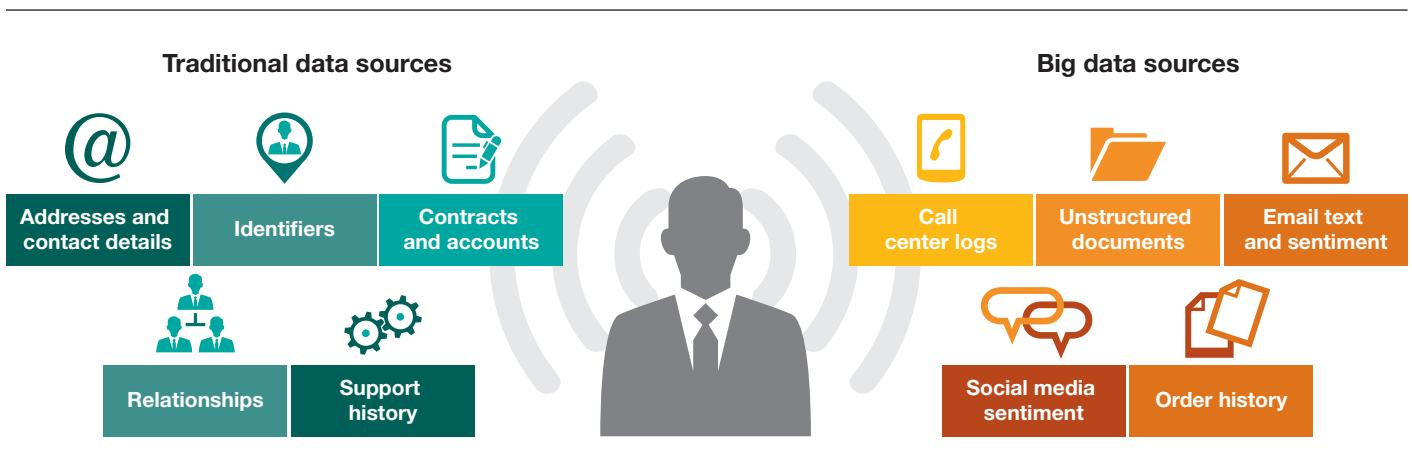

Figure 1. InfoSphere Big Match for Hadoop offers an efficient approach to finding insight in both traditional and big data sources.

The solution, in combination with the Text Analytics capabilities in IBM BigInsights® for Apache Hadoop, addresses structured and unstructured data using a combination of text analytics and matching. It provides organizations with the ability to parse out entity information buried in unstructured data (for example, Twitter handle information) and match these to entities such as customers from other sources natively on Hadoop.

Further, InfoSphere Big Match for Hadoop is flexible enough to match entities based on configurable thresholds most appropriate to their use. Organizations can employ tighter thresholds for matching where data sets are known to have stringent governance, and where a high degree of confidence in the match is necessary. Alternatively, matching thresholds can be relaxed where data sets may lack governance (for example, social media feeds) and the requirement for matching accuracy is lower.

Handle tremendous data volumes

As the term "big data" implies, organizations are struggling to manage large volumes of data that exist within and outside of their systems and infrastructure. In many cases, Hadoop allows organizations to gain control of the growing amounts of data that cannot be efficiently or economically managed by traditional systems regardless of volume size.

Although Hadoop enables organizations to create data lakes, the requirements of preparing customer data for analytics can extend beyond Hadoop's native features. Matching is the most critical of these requirements, as it creates both accuracy (identifying duplication) and completeness (linking fragmented data) in analytic customer insight.

It's not easy to acquire this capability. Traditional tools often cannot be effectively employed in big data environments—they are either unavailable on such platforms or are overwhelmed by the sheer volume of data, taxing hardware and impeding results. A slowdown in customer analytics translates into delayed customer offers and a lack of true insight into markets or sentiment, which can dramatically impact the business.

InfoSphere Big Match for Hadoop can rapidly churn through massive data sets numbering in the billions of records, linking customer information with intelligence and efficiency. Using a combination of distributed probabilistic matching technology, big data accelerators and text analytics, the solution extracts relevant information and helps connect customer identities at the speed of business. Within data lake architectures, InfoSphere Big Match for Hadoop provides a consistent customer dimension for all data analytics. Without the need for pre-processing, organizations can better understand the value of a transaction relative to the customer data before deciding how and where to further invest. InfoSphere Big Match for Hadoop complements existing master data management investments by offloading analytical workloads, thereby optimizing performance and business response times.

For more information

To learn more about InfoSphere Big Match for Hadoop, contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/software/products/en/infosphere-big-match-for-hadoop

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing

IBM

[1] "Analytics: The real-world use of big data." A collaborative research study
by the IBM Institute for Business Value and the Saïd Business School at
the University of Oxford. October 2012. **ibm.com**/services/us/gbs/
thoughtleadership/ibv-big-data-at-work.html

Please Recycle