INFORMATICA®
The Data Integration Company™

Big Data Unleashed

Turning Big Data into Big Opportunities with the Informatica 9.1 Platform

WHITE PAPER

This edition published May 2011

Table of Contents

Executive Summary

Big Data—from multi-petabyte data warehouses to social media data, from cloud-based applications to sensors and mobile devices, from e-commerce processing to geospatial information—has arrived. Whether Big Data becomes an organization's greatest asset or one of its gravest liabilities depends on the strategies and solutions put in place in the near term to deal with epic growth in data volumes, complexity, diversity, and velocity.

Informatica® 9.1 is expressly engineered to help your organization turn Big Data into big opportunities. The latest release of the Informatica Platform empowers organizations to combine fast-growing transaction data with vast volumes of new interaction data to achieve insights and strategic advantage not possible with any other solution.

This white paper examines the challenges and opportunities with Big Data, outlines real-world use cases of organizations using Informatica to transform Big Data into business benefits, and details groundbreaking capabilities in Informatica 9.1 to empower the data-centric enterprise. We will focus on four key themes:

- Big Data integration
- Authoritative and trustworthy data
- Self service
- Adaptive data services

The Big Data Megatrend

Remember when a 1 TB data warehouse was considered huge? Now you can buy a 1 TB storage device for less than $100 at your local retailer, while many data warehouses have broken through the petabyte barrier.

Speaking of petabytes, quick—how many petabytes in a zettabyte? Until just recently, the term zettabyte was used only in a theoretical sense. Now experts estimate that the world's data has exceeded the ZB threshold—more than 1 million petabytes.

Welcome to the era of Big Data. Over the next several years, the volume of data that enterprises generate, consume, store, and have access to will increase exponentially. Some have called Big Data "the Industrial Revolution of data"—a historic and game-changing expansion of the role that information plays in business, government, and consumer realms.

But growing data volumes are only half the Big Data equation. Big Data also introduces large-scale increases in data diversity, complexity, and velocity. Already many organizations confront a relentless onslaught of Big Data—data in dozens of forms, from hundreds of sources, streaming in real time on a global scale. Some believe that Big Data will dwarf all that has come before it.

The research firm IDC summed up the Big Data phenomenon this way: "You ain't seen nothing yet."[1]

[1] IDC, "The Digital Universe Decade—Are You Ready?" May 2010.

The "Disruptive Force" of Big Data

Is your organization developing strategies and solutions not only to deal with the growth of Big Data but also to turn Big Data into big opportunities? Are you prepared to integrate and analyze Big Data to improve decision making and aggressively pursue growth? Even streamline business processes by injecting the information into operational flows? With the right platform and processes, the business potential of Big Data can be unleashed to power the data-centric enterprise. Organizations at the leading edge of Big Data use will achieve competitive advantages in such areas as customer management, sales performance, operational efficiency, and business agility.

Yet today, few organizations are equipped to access and process the full scope of Big Data. IT departments may be feeling the brunt of escalating expectations of information delivery to executives and managers, front-line staff, customers, or partners. They lack the flexible, scalable data and information infrastructures required to exploit Big Data for critical insights that translate into advantage.

Inability to cope with the Big Data onslaught is not just an opportunity lost. It introduces a substantial risk. Rising data volumes can obscure visibility into opportunities and threats. Data complexity can compromise compliance. Ceaseless streams of real-time data from multiple channels can degrade customer sales and service.

> "Big Data is a disruptive force and an immediate problem that is already affecting traditional understanding and business models..... It represents a disruption in current trends, and represents a huge opportunity that public sector, business, and IT leaders cannot afford to ignore ,"
>
> Gartner[2]

Big Data Defined

What is Big Data? Big Data means all data, including both transaction and interaction data, in sets whose size or complexity exceeds the ability of commonly used technologies to capture, manage, and process at a reasonable cost and timeframe. In fact, Big Data is the confluence of three major technology trends as Figure 1 illustrates:

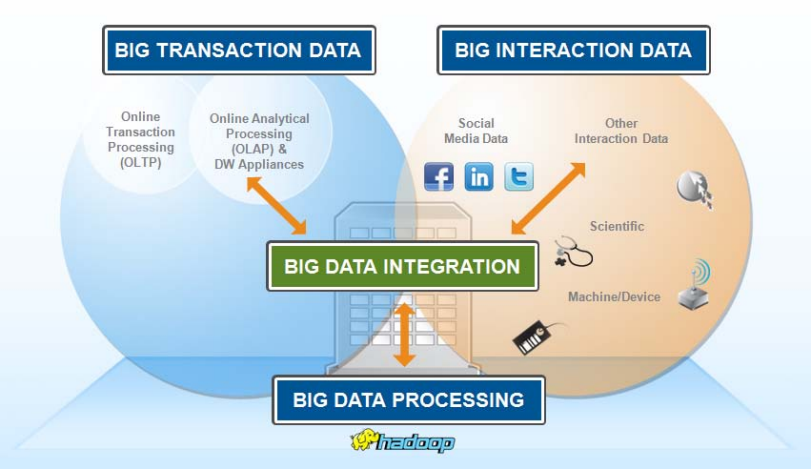

Figure 1. Big Data is the confluence of three trends consisting of Big Transaction Data, Big Interaction Data, and Big Data processing.

[2] Gartner, "CEO Advisory: 'Big Data' Equals Big Opportunity," March 31, 2011.

- **Big Transaction Data:** Traditional relational data continues to grow in on-line transactional processing (OLTP) and analytic systems, from ERP applications to data warehouse appliances, along with unstructured and semistructured information. The landscape is complicated as enterprises move more data and business processes to public and private clouds.
- **Big Interaction Data:** This emerging force consists of social media data from Facebook, Twitter, LinkedIn, and other sources. It includes call detail records (CDRs), device and sensor information, GPS and geolocational mapping data, large image files through Manage File Transfer, Web text and clickstream data, scientific information, emails, and more.
- **Big Data Processing:** The rise of Big Data has given rise to frameworks geared for data-intensive processing such as the open-source Apache Hadoop, running on a cluster of commodity hardware. The challenge for enterprises is to get data into and out of Hadoop rapidly, reliably, and cost-effectively.

How Big Is Big?

While experts agree that Big Data is big, exactly how big is a matter of debate. IDC forecasts a roughly 50 percent annual growth rate for what it calls the world's "digital universe," more than 70 percent of which IDC estimates is generated by consumers and over 20 percent by enterprises. Between 2009 and 2020, the digital universe will swell by a factor of 44 to 35 zettabytes, or 35 million petabytes, IDC predicts.[3]

Big Data is estimated to be far bigger by computer scientists at the University of California at San Diego. In an April 2011 report, UC San Diego estimated the world's enterprise servers processed 9.57 ZB of data in 2008—not counting 3.6 ZB it calculated was generated by American households.[4] UC San Diego's figure for enterprise-only information is more than 10 times greater than IDC's all-inclusive approximation of both enterprise and consumer data, but the bottom line is it is big either way. Putting its estimate in perspective, UC San Diego said 9.75ZB in paperback book form would stretch from Earth to Neptune and back 20 times.

Big Data in the Enterprise

Within an individual enterprise, no specific size defines Big Data, but generally it falls within a few dozen terabytes to many petabytes. A survey by Unisphere Research of 531 members of the Independent Oracle Users Group found that enterprise data volumes are growing rapidly at 9 of 10 companies, with 16 percent experiencing increases of 50 percent or more each year.[5]

Enterprises already are feeling the impact of uncontrolled data growth on performance. For example, an Informatica-sponsored survey by Unisphere Research found that 87 percent of respondents blamed increasing data volumes for enterprise application performance problems.[6]

[3] IDC, "The Digital Universe Decade—Are You Ready?" Sponsored by EMC, May 2010.

[4] University of California at San Diego, "How Much Information? 2010 Report on Enterprise Server Information," April 2011.

[5] Unisphere Research, "Keeping Up with Ever-Expanding Enterprise Data," October 2010.

[6] Unisphere Research, "Information, Unplugged: 2009 OAUG ResearchLine Survey on Enterprise Application Information Lifecycle Management," April 2010.

Seizing the Big Data Business Opportunity

What can your organization do with Big Data? How can you take advantage of its big opportunities? How can you avoid its risks? An increasing number of organizations tackling Big Data are deploying more advanced massively parallel processing (MPP) databases, Hadoop distributed file systems, MapReduce algorithms, cloud computing, and archival storage. The linchpin is data integration. It's crucial for organizations to enable business to access all data so they can apply it across Big Data infrastructures.

Data integration enables your organization to hit the Big Data sweet spot—combining traditional transaction data with new interaction data to generate insights and value otherwise unachievable. A prime example is enriching customer profiles with likes and dislikes culled from social media to improve targeted marketing. Without data integration, Big Data amounts to lots of Big Data silos.

As Big Data comes into focus, it's capturing the attention of CIOs, VPs of information management (IM), enterprise architects, line-of-business owners, and business executives who recognize the vital role that data plays in performance. Indeed, CEOs named "data-driven decision-making" as the technology contribution "delivering the most strategic value to the business," according to a 2011 Gartner survey of CEOs and senior executives.[7] Big Data is relevant to virtually every industry:

- **Consumer industries:** From retail to travel and hospitality, organizations can capture Facebook posts, Twitter tweets, YouTube videos, blog commentary, and other social media content to better understand, sell to, and service customers, manage brand reputation, and leverage word-of-mouth marketing.
- **Financial services:** Banks, insurers, brokerages, and diversified financial services companies are looking to Big Data integration and analytics to better attract and retain customers and enable targeted cross-sell, as well as strengthen fraud detection, risk management, and compliance by applying analytics to Big Data.
- **Public sector:** Federal Networking and Information Technology Research and Development (NITRD) working group announced the Designing a Digital Future report. The report declared that "every federal agency needs a Big Data strategy," supporting science, medicine, commerce, national security, and other areas; state and local agencies are coping with similar increases in data volumes in such diverse areas as environmental reviews, counter terrorism and constituent relations.
- **Manufacturing and supply chain:** Managing large real-time flows of radio frequency identification (RFID) data can help companies optimize logistics, inventory, and production while swiftly pinpointing manufacturing defects; GPS and mapping data can streamline supply chain efficiency.
- **E-commerce:** Harnessing enormous quantities of B2B and B2C clickstream, text, and image data and integrating them with transactional data (such as customer profiles) can improve e-commerce efficiency and precision while enabling a seamless customer experience across multiple channels.
- **Healthcare:** The industry's transition to electronic medical records and sharing of medical research data among entities is generating vast data volumes and posing acute data management challenges; biotech and pharmaceutical firms are focusing on Big Data in such areas as genomic research and drug discovery.
- **Telecommunications:** Ceaseless streams of CDRs, text messages, and mobile Web access both jeopardize telco profitability and offer opportunities for network optimization. Firms are looking to Big Data for insights to tune product and service delivery to fast-changing customer demands using social network analysis and influence maps.

[7] Gartner, "CEO Advisory: 'Big Data' Equals Big Opportunity," March 31, 2011.

Overcoming the Obstacles of Existing Data Infrastructures

Traditional approaches to managing data are insufficient to deliver the value of business insight from Big Data sources. The growth of Big Data stands to exacerbate pain points that many enterprises suffer in their information management practices:

- **Lack of business/IT agility.** The IM organization is perceived as too slow and too expensive in delivering solutions that the business needs for data-driven initiatives and decision making.
- **Compromised business performance.** IM constantly deals with complaints from business users about the timeliness, reliability, and accuracy of data while lacking standards to ensure enterprise-wide data quality.
- **Overreliance on IM.** The business has limited abilities to directly access the information it needs, requiring time-consuming involvement of IM and introducing delays into critical business processes.
- **High costs and complexity.** The enterprise suffers escalating costs due to data growth and application sprawl, as well as degradation of systems performance, leaving it poorly positioned for the Big Data onslaught.
- **Delays and IT re-engineering.** Costly architectural rework is necessary when requirements change even slightly, with little reuse of data integration logic across projects and groups.
- **Lost customer opportunities.** Sales and service lack a complete view of the customer, undercutting revenue generation and missing opportunities to leverage behavioral and social media data.

Of these problems, addressing the limitations of existing CRM systems and exploiting Big Data from social media sources to attract and retain customers and improve cross-sell effectiveness are of keen interest to executives. Organizations are transitioning to CRM 2.0, which depends fundamentally on a complete and accurate customer view from large and diverse data sources. According to a Gartner report:

"The Gartner 2011 CEO and Senior Business Executive Survey revealed that 'better data' and a 'better understanding of the customer' were high priorities for business leaders. Big Data has the potential to deliver both, but only for organizations that invest the necessary resources in business processes, management actions and technology to master this trend."[8]

[8] Gartner, "CEO Advisory: 'Big Data' Equals Big Opportunity," March 31, 2011.

Informatica 9.1: Built for Big Data

The latest release of the Informatica Platform, Informatica 9.1, was developed with the express purpose of turning Big Data challenges into big opportunities. Building on 18 years of data integration leadership, Informatica 9.1 is engineered to empower the data-centric enterprise to unleash the business potential of Big Data in four areas:

- Big Data integration to gain business value from Big Data
- Authoritative and trustworthy data to increase business insight and consistency by delivering trusted data for all purposes
- Self-service to empower all users to obtain relevant information while IT remains in control
- Adaptive data services to deliver relevant data adapted to the business needs of all projects

The next section outlines capabilities in Informatica 9.1 and how it enables your organization to tackle Big Data opportunities.

Big Data Integration

Informatica 9.1 delivers innovations and new features in the three areas of Big Data integration:

Connectivity to Big Transaction Data. Informatica 9.1 provides access to high volumes of transaction data, up to a petabyte in scale, with native connectivity to OLTP and on-line analytical processing (OLAP) data stores. A new relational/data warehouse appliance package available in Informatica 9.1 extends this connectivity to solutions purpose-built for Big Data.

- Maximize the availability and performance of large-scale transaction data from any source
- Reduce the cost and risk of managing connectivity with a single platform supporting all database and processing types
- Uncover new areas for growth and efficiency by leveraging transaction data in a scalable, cost-effective way

Connectivity to Big Interaction Data. Access new sources such as social media data on Facebook, Twitter, LinkedIn, and other media with new social media connectors available in Informatica 9.1. Extend your data reach into emerging data sets of value in your industry, including devices and sensors, CDRs, large image files, or healthcare-related information for biotech, pharmaceutical, and medical companies.

- Gain new insights into customer relationships and influences enabled by social media data
- Access and integrate other types of Big Interaction Data and combine it with transaction data to sharpen insights and identify new opportunities
- Reduce the time, cost, and risk of incorporating new data sets and making them available to enterprise users

"LinkShare has been the pioneer in turning massive amounts of campaign and advertising data into business value for our customers and partners. Now Big Data is making pervasive impact to every part of LinkShare's online pay-per-action marketing network. By leveraging Informatica 9.1 for Big Data, Adaptive Data Services, and Self-Service, LinkShare will be able to roll out new products faster, meet customer demands for reporting and analysis more effectively, increase reuse of integration logic, and consequently accelerate our growth."

David Ramos
Director of Business Intelligence and Analytics
LinkShare

Big Data Processing. New connectivity in Informatica 9.1 enables IT to load data from any source into Hadoop and extract data from Hadoop for delivery to any target. The connectivity also allows the application of Informatica data quality, data profiling, and other techniques to data in Hadoop. These capabilities open new possibilities for enterprises combining transaction and interaction data either inside or outside of Hadoop.

- Confidently deploy the Hadoop platform for Big Data processing with seamless source-and-target data integration
- Integrate insights from Hadoop Big Data analytics into traditional enterprise systems to improve business processes and decision making
- Leverage petabyte-scale performance to process large data sets of virtually any type and origin

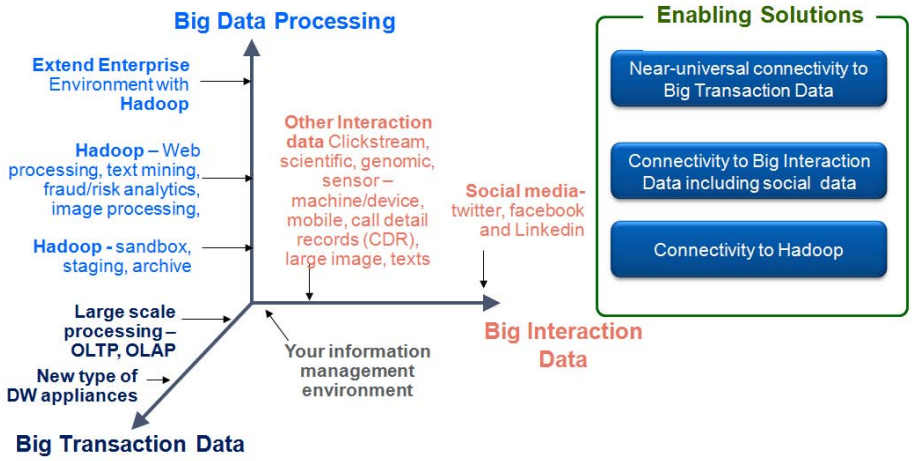

Figure 2. Big Data integration involves the ability to harness Big Transaction Data, Big interaction Data, and Big Data processing.

Big Data Integration in Action

Every new data source is a new business opportunity. Whether it's social media data posted by your Facebook fans, sensor-based RFID information in your product supply chain, or the enterprise applications of a newly acquired company, your ability to harness this information bears directly on your bottom line.

Unleashing the potential of Big Data requires the ability to access and integrate information of any scale, from any source. In many cases, it means combining interaction data with transaction data to enable insights not possible any other way. One example is using social media data to drive revenue by attracting and retaining customers.

With 50 million tweets on Twitter and 60 million updates on Facebook daily and going up, consumers are sharing insights into what they like and don't like. Suppose your company could learn from a Facebook fan that her son is looking for colleges, she's shopping for a new car, and she likes Caribbean cruises? That's invaluable intelligence for targeted marketing and customer loyalty projects.

Moreover, organizations are learning that social media fans double as salespeople, triggering viral word-of-mouth purchasing among friends and friends of friends in a broad sphere of influence that makes these customers uniquely valuable. The opposite can be true with cranky customers damage business reputation when they complain about poor experiences with companies. Yet with just 20 percent of social media data estimated to be relevant for marketing, organizations need a high-performance platform capable of parsing select content from dynamic data streams.

With Informatica 9.1, your organization can harness social media data to enrich customer profiles in CRM applications with customer likes, dislikes, interests, business and household information, and other details. Support for Hadoop gives you data interoperability between the distributed processing framework and your transactional systems, with flexibility for bidirectional data movement to meet your business objectives.

Authoritative and Trustworthy Data

Informatica 9.1 supplies master data management (MDM) and data quality technologies to enable your organization to achieve better business outcomes by delivering authoritative, trusted data to business processes, applications, and analytics, regardless of the diversity or scope of Big Data.

Single platform for all MDM architectural styles and data domains. Universal MDM capabilities in Informatica 9.1 enable your organization to manage, consolidate, and reconcile all master data, no matter its type or location, in a single, unified solution. Universal MDM is defined by four characteristics:

- **Multi-domain:** Master data on customers, suppliers, products, assets, locations, can be managed, consolidated, and accessed.
- **Multi-style:** A flexible solution may be used in any style: registry, analytical, transactional, or co-existence.
- **Multi-deployment:** The solution may be used as a single-instance hub, or in federated, cloud, or service architectures.
- **Multi-use:** The MDM solution interoperates seamlessly with data integration and data quality technologies as part of a single platform.

Universal MDM eliminates the risk of standalone, single MDM instances—in effect, a set of data silos meant to solve problems with other data silos.

- Flexibly adapt to different data architectures and changing business needs
- Start small in a single domain and extend the solution to other enterprise domains, using any style
- Cost-effectively reuse skill sets and data logic by repurposing the MDM solution

"No data is discarded anymore! U.S. xPress leverages a large scale of transaction data and a diversity of interaction data, now extended to perform big data processing like Hadoop with Informatica 9.1. We assess driver performance with image files and pick up customer behaviors from texts by customer service reps. U.S. xPress saved millions of dollars per year by reducing fuels and optimizing routes augmenting our enterprise data with sensor, meter, RFID tags, and geospatial data."

Tim Leonard
Chief Technology Officer
U.S. xPress

Reusable data quality policies across all project types. Interoperability among the MDM, data quality, and data integration capabilities in Informatica 9.1 ensures that data quality rules can be reused and applied to all data throughout an implementation lifecycle, across both MDM and data integration projects (see Figure 3).

- Seamlessly and efficiently apply data quality rules regardless of project type, improving data accuracy
- Maximize reuse of skills and resources while increasing ROI on existing investments
- Centrally author, implement, and maintain data quality rules within source applications and propagate downstream

Proactive data quality assurance. Informatica 9.1 delivers technology that enables both business and IT users to proactively monitor and profile data as it becomes available, from internal applications or external Big Data sources. You can continuously check for completeness, conformity, and anomalies and receive alerts via multiple channels when data quality issues are found.

- Receive "early warnings" and proactively identify and correct data quality problems before they happen
- Prevent data quality problems from affecting downstream applications and business processes
- Shorten testing cycles by as much as 80 percent

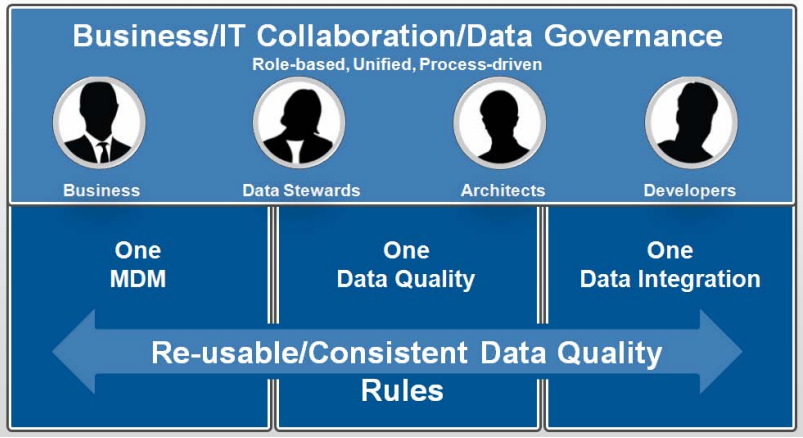

Figure 3. Informatica 9.1 unifies data integration, data quality, and MDM to deliver authoritative and trustworthy data.

Putting Authoritative and Trustworthy Data to Work

The diversity and complexity of Big Data can worsen the data quality problems that exist in many organizations. Standalone, ad hoc data quality tools are ill equipped to handle large-scale streams from multiple sources and cannot generate the reliable, accurate data that enterprises need. Bad data inevitably means bad business. In fact, according to a CIO Insight report, 46 percent of survey respondents say they've made an inaccurate business decision based on bad or outdated data.[9]

MDM and data quality are prerequisites for making the most of the Big Data opportunity. Here are two examples:

[9] CIO Insight, "The Big Data Conundrum," November 9, 2010

- **Using social media data to attract and retain customers.** For some organizations, tapping social media data to enrich customer profiles can be putting the cart before the horse. Many companies lack a single, complete view of their customers, ranging from reliable and consistent names and contact information to the products and services in place. Customer data is often fragmented across CRM, ERP, marketing automation, service, and other applications. Informatica 9.1 MDM and data quality enable you to build a complete customer profile from multiple sources. With that authoritative view in place, you're poised to augment it with the intelligence you glean from social media.
- **Data-driven response to business issues.** Let's say you're a Fortune 500 manufacturer and a supplier informs you that a part it sold you is faulty and needs to be replaced. You need answers fast to critical questions: In which products did we use the faulty part? Which customers bought those products and where are they? Do we have substitute parts in stock? Do we have an alternate supplier?

  But the answers are sprawled across multiple domains of your enterprise—your procurement system, CRM, inventory, ERP, maybe others in multiple countries. How can you respond swiftly and precisely to a problem that could escalate into a business crisis? Business issues often span multiple domains, exerting a domino effect across the enterprise and confounding an easy solution. Addressing them depends on seamlessly orchestrating interdependent processes—and the data that drives them.

  With the universal MDM capabilities in Informatica 9.1, our manufacturer could quickly locate reliable, authoritative master data to answer its pressing business questions, regardless of where the data resided or whether multiple MDM styles and deployments were in place.

Self-Service

Big Data's value is limited if the business depends on IT to deliver it. Informatica 9.1 enables your organization to go beyond business/IT collaboration to empower business analysts, data stewards, and project owners to do more themselves without IT involvement with the following capabilities illustrated in Figure 4.

Self-service data integration for analysts and data stewards. In a collaborative framework with IT, business users can implement rules for data integration and quality in role-based, nontechnical interfaces through Informatica 9.1. Analysts and data stewards can assume a greater role in defining specifications, promoting a better understanding of the data, and improving productivity for business and IT.

- Empower business users to access data based on business terms and semantic metadata
- Accelerate data integration projects through reuse, automation, and collaboration
- Minimize errors and ensure consistency by accurately translating business requirements into data integration mappings and quality rules

Application-aware accelerators for project owners.

Informatica 9.1 empowers project owners to rapidly understand and access data for data warehousing, data migration, test data management, and other projects. Project owners can source business entities within applications instead of specifying individual tables that require deep knowledge of the data models and relational schemas.

- Reduce data integration project delivery time
- Ensure data is complete and maintains referential integrity
- Adapt to meet business-specific and compliance requirements

> "Traditional methods of using spreadsheets to manage specifications can be inefficient and error prone. We are excited about the self-service data integration capabilities in Informatica 9.1 to help us deliver projects faster. The Informatica Data Integration Analyst option improves business-IT collaboration and eliminates errors by empowering the business to interactively define specifications and automatically generate mappings for deployment to production."
>
> Marco Eugênio Carvalho
> Integration Architect
> Vale

Point-of-use data and context for business users. Informatica 9.1 gives business users authoritative, trustworthy data in the business applications they use every day, rather than accessing data from multiple systems. Users can also view and leverage context such as hierarchical and relationship information, business glossary, and data lineage important in a range of processes.

- Increase the accuracy and timeliness of business decisions
- Improve business productivity and operational efficiency
- Minimize dependence on limited IT resources, liberating them to pursue value-added projects

Proactive data quality assurance. Addressing the velocity and complexity issues of Big Data, Informatica 9.1 introduces complex event processing (CEP) technology into data quality and integration monitoring to alert business users and IT of issues in real time. For instance, it will notify an analyst if a data quality key performance indicator exceeds a threshold, or if integration processes differ from the norm by a predefined percentage.

- Enable business users to define monitoring criteria by using prebuilt templates
- Alert business users on data quality and integration issues as they arise
- Identify and correct problems before they impact performance and operational systems

Figure 4. Self-service empowers all users to access critical information and contribute to data integration and quality while IT remains in control.

Empowering the Business with Self-Service

Business and IT alike have long been handicapped by a lack of collaboration and data self-service for business users. The business lacks the reliable information it needs, while productivity suffers as users waste time hunting for information. The advent of Big Data and the additional complexity of data landscapes promise to worsen the problem. Informatica 9.1 equips business users with self-service and promotes greater business/IT collaboration.

- **Speeding and strengthening business effectiveness.** Informatica 9.1 makes "MDM-aware" everyday business applications such as Salesforce.com, Oracle, Siebel, SAP for CRM, ERP, and others by presenting reconciled master data directly within those applications. For example, Informatica's MDM solution will advise a salesperson creating a new account for "John Jones" that a customer named Jonathan Jones, with the same address, already exists. Through the Salesforce interface, the user can access complete, reliable customer information that Informatica MDM has consolidated from disparate applications.

  She can see the products and services that John has in place and that he follows her company's Twitter tweets and is a Facebook fan. She has visibility into his household and business relationships and can make relevant cross-sell offers. In both B2B and B2C scenarios, MDM-aware applications spare the sales force from hunting for data or engaging IT while substantially increasing productivity.

- **Giving business users a hands-on role in data integration and quality.** Long delays and high costs are typical when the business attempts to communicate data specifications to IT in spreadsheets. Part of the problem has been the lack of tools that promote business/IT collaboration and make data integration and quality accessible to the business user.

  As Big Data unfolds, Informatica 9.1 gives analysts and data stewards a hands-on role. Let's say your company has acquired a competitor and needs to migrate and merge new Big Data into your operational systems. A data steward can browse a data quality scorecard and identify anomalies in how certain customers were identified and share a sample specification with IT. Once validated, the steward can propagate the specification across affected applications. A role-based interface also enables the steward to view data integration logic in semantic terms and create data integration mappings that can be readily understood and reused by other business users or IT.

We are excited about Informatica 9.1 as it is truly the industry's most advanced data integration platform. With Informatica 9.1, we can access any data including interactional data, process the data anywhere, deliver the data at any latency to any consuming application, and maximize productivity with self-service and business-IT collaboration. We can reuse data services across all projects and mitigate risk by proactively ensuring that the data use is consistent and accurate. We can future-proof our business by leveraging flexible deployment options that scale per our needs. Informatica Data Services offers the best ROI for enterprises like ours' that are looking to become even more agile.

Rob Meyers
Manager of Business Intelligence Architecture
Healthnow

Adaptive Data Services. Informatica 9.1 extends the services-centric platform with adaptive data services to enable your enterprise to deliver relevant data adapted to the business needs of all projects.

Multi-protocol data provisioning. Informatica 9.1 helps you provision data services as a SQL endpoint through ODBC or JDBC, as a Web service, or to Informatica PowerCenter® as a batch process. It serves as a data virtualization layer for SOA implementations. This approach helps you readily combine real-time data with historical data stored in data warehousing and other analytical systems.

- Improve developer productivity through seamlessly reusing data services without any rework
- Reduce project time to delivery through data virtualization
- Ensure consistency across projects

Integrated data quality. Informatica 9.1 allows you to ensure that all data provisioned via data services is of high quality. It enforces data quality rules at the point of access—both read and write. You can reuse the library of templates and data quality rules when deploying the integrated data quality capability.

- Ensure the business has trustworthy information throughout the enterprise
- Minimize impact of bad data by pushing data quality as business sees it
- Deploy data quality in real-time without the need for pre or post-processing or staging

Policy-driven enforcement. Informatica 9.1 helps you deploy reusable data services for data access, data freshness, data quality, data retention, and data privacy. These policies conform with standards and comply with regulations adapted to the needs of the business and associated data requirements. It leverages caching rules to specify how often data needs to be refreshed and whether data is to be delivered in real time, near real time, or batch. It also deploys data retention policies (e.g., specify when to archive dormant data and how long to retain archived data).

- Reduce the risk of noncompliance with policies, regulations, and SLAs
- Define and enforce policies consistently across all applications for data governance
- Help lower the costs of complying with multiple regulations on a single platform

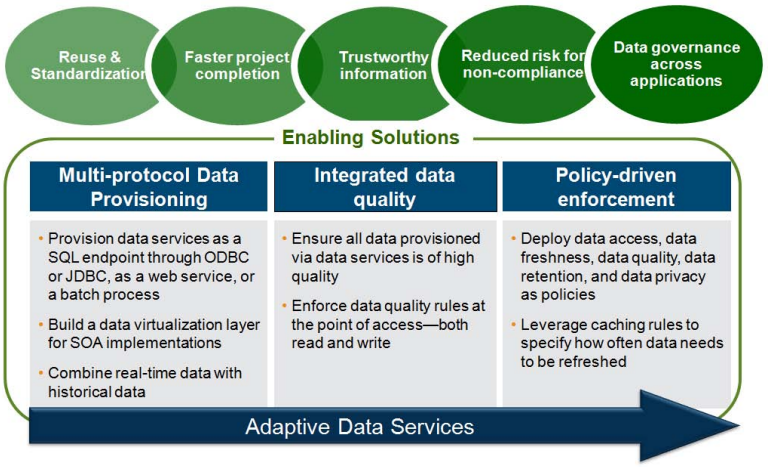

Figure 5. Adaptive Data Services delivers relevant data adapted to the business needs of all projects.

Virtualizing Your Data Access with Adaptive Data Services

The data virtualization and federation capabilities of Informatica 9.1 supply a powerful and elegant solution to dealing with the increased complexity of Big Data. Enterprise architects and developers can swiftly design, implement, and reuse a range of data services in substantially less time and at lower cost than traditional approaches.

This innovative technology enables you to combine data from heterogeneous sources as a single virtual view that acts as one data source to consuming applications.]With support for all key standards, data services give you the flexibility to expose data integration logic as reusable services and schedule complex data transformations from any source either on demand or at scheduled intervals.

Integrated data profiling services let you assess the structure and quality of data at any stage in its lifecycle, while integrated data quality services enable you to ensure that trustworthy and authoritative data is delivered to the business. Data services may be seamlessly reused as federated views for reporting and business intelligence solutions that access the data directly from source systems, as part of transformation processes for data warehousing, or as services of real-time data integration.

Customer Use Cases: Big Data in Action

Informatica customers are pioneering initiatives that harness the power of Big Data.

Online Marketing Leader for Cost Reduction

This online affiliate marketing company is one of the world's largest Internet service companies, providing leading services in e-commerce, portal and media, travel, financial services, and professional sports. It provides an online marketing platform that integrates high-quality direct response digital media, services, and technology. It enables clients to develop cost-efficient pay-per-action affiliate, search, and lead generation campaigns that acquire new customers, increase revenue, and drive results.

This Informatica customer has a strong incentive to position the right ad to the right person at the right time to maximize revenues. That alone is a huge optimization problem. It also has an SLA with its customers that mandates an up-to-date, accurate invoice to be delivered within the last five minutes of activity. The challenges involved a massive volume of transaction data coming from more than 500 data sources, concerning over 300 million transactions a day, and serving information to 300,000 users in real time. It used Informatica to make it all possible and has seen a 3x reduction in cost and has delivered projects up to five times faster using Informatica.

Large Department Store Retailer for Customer Centricity

One of the leading fashion specialty retailers serves its customers through local department stores, online, and its catalog business. The company is known for its differentiated services for its clientele. After some analysis, the retailer decided to stop giving free makeup services and cosmetic samples because managers realized that customers receiving these freebies were not buying more cosmetics. The retailer expected that cosmetics sales would remain the same once its giveaway program ended, but instead experienced a decline in cosmetics sales.

Through research, including harvesting social media information from Twitter and Facebook, it started to better understand the influence model for cosmetics. It learned that it has two types of valuable customers that must be retained—high spenders and high influencers. Customers receiving a free makeup session weren't necessarily buying cosmetics, but their word of mouth was prompting purchases by friends and friends of friends. This was a perfect marriage of transaction data and interaction data to come up with nonobvious answers to a business challenge. By using Informatica, this retailer enriched its customer master data with social media data and made its services more targeted. It increased profits through treating those high-influence customers with the right offers.

Transportation Mobile Intelligence Pioneer for Improved Processes

This trucking company's vision is to become a leader in transportation mobile intelligence with next-generation onboard communication systems and thousands of drivers and tens of thousands of trucks and other vehicles. It is focused on using mobile intelligence to optimize operations by tracking and getting insights from all the movements of its assets, trucks, drivers, operation managers, partners, and customers. It faced several challenges. The company had no means of determining where trucks had stopped or for how long. Money was wasted on engine idle time, trucks weren't being used as efficiently as possible, and customer service was suffering.

As part of its "No Data Left Behind" program, this company started collecting 900 data elements from trucking systems, several times a second, effectively tracking every piece of the data imaginable—sensor data for tire and gas usage, engine operation, and geospatial data to track the fleet. It even scraped data from blogs by truckers complaining about the system so it could make improvements. A Hadoop-based Big Data processing environment and Informatica were key parts of the enterprise information management system that enabled the company to optimize fleet use, reduce emissions, and meet environmental commitments. This helped it save millions of dollars annually and helped us all "go green."

Conclusion

One thing is certain about Big Data. It will only get bigger. The data landscape of tomorrow will be fertile with opportunities to improve performance across multiple domains, yet riddled with the pitfalls posed by rising data volumes, complexity, diversity, and velocity.

Big Data is here today, and so is the data integration technology that your organization can use to meet this megatrend head-on. By harnessing and combining large-scale transactional data with new interaction data and taking advantage of data-intensive frameworks, your organization will be prepared to realize the big opportunities of Big Data and to become a data-centric enterprise.

Learn More

Learn more about the Informatica Platform. Visit us at www.informatica.com or call +1 650-385-5000 (1-800-653-3871 in the U.S.).

About Informatica

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world gain a competitive advantage in today's global information economy with timely, relevant and trustworthy data for their top business imperatives. More than 4,200 enterprises worldwide rely on Informatica to access, integrate and trust their information assets held in the traditional enterprise, off premise and in the Cloud.

INFORMATICA®
The Data Integration Company™

Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

1601 (05/24/2011)