

# The Informatica Platform for the U.S. Intelligence Community

The Data Integration Leader Supports the Charge to Collect Intelligence



This document contains Confidential, Proprietary and Trade Secret Information (“Confidential Information”) of Informatica Corporation and may not be copied, distributed, duplicated, or otherwise reproduced in any manner without the prior written consent of Informatica.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Informatica does not accept responsibility for any kind of loss resulting from the use of information contained in this document. The information contained in this document is subject to change without notice.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Informatica software product—as well as the timing of any such release or upgrade—is at the sole discretion of Informatica.

Protected by one or more of the following U.S. Patents: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; or by the following pending U.S. Patents: 09/644,280; 10/966,046; 10/727,700.

This edition published January 2013

## Table of Contents

<b>Executive Summary</b> .....	<b>2</b>
<b>Unique Needs of the U.S. Intelligence Community</b> .....	<b>3</b>
Detecting Suspicious Activity by Proactively Identifying Risks, Anomalies, and Unusual Activities .....	4
Delivering Unprecedented Insight Through Proactive Monitoring and Alerting .....	4
Establishing a Data Governance Plan to Use Data Efficiently Across the Organization .....	4
<b>The Informatica Platform</b> .....	<b>5</b>
Unified .....	5
Open .....	6
Economical .....	6
Comprehensive .....	7
<b>The Informatica Platform in Action</b> .....	<b>9</b>
Case Study 1: Onboard Data Faster .....	9
Case Study 2: Optimize Enterprise Search & Discovery .....	10
<b>Conclusion</b> .....	<b>11</b>

“Data is created constantly, often unknowingly and without permission. The number of data points that can be collected is virtually limitless—presenting, of course, both enormous intelligence opportunities and equally large counterintelligence challenges.”

General David Petraeus  
In-Q-Tel CEO Summit,  
March 2012

## Executive Summary

U.S. intelligence agencies are tasked with the ongoing mission of safeguarding against acts of terrorism and cyber warfare. They must constantly reinvent methods of garnering new insights and detecting ever-evolving threats.

A fundamental imperative to this community is the ability to identify relevant information and the potential relationships of that information to other data points and to share information within and among organizations. Critical to accomplishing these tasks are the collection, analysis, and dissemination of key transactions, events, and communications. But in carrying out their missions, intelligence agencies face numerous challenges to gathering, categorizing, handling, storing, and securing this data.

Today, the intelligence community is being bombarded with new data from multiple sources at a nonstop pace. This critical function is further complicated by the sheer volume of input that is coming from both traditional and untraditional sources with countless formats.

As organizations strive to keep pace, analysts in the intelligence community increasingly need to proactively interpret the complexities presented to them with unparalleled precision and sophistication. They require cost-effective, secure, and flexible software to handle the most complex data demands and to meet stringent retention and privacy mandates.

This white paper discusses the Informatica® Platform and explains how this enterprise data integration platform’s open, unified, economical, and comprehensive features meet the needs of intelligence analysts. Included are case studies illustrating how the Informatica Platform has been used successfully in intelligence agency environments.

# Unique Needs of the U.S. Intelligence Community

U.S. intelligence agencies grapple with the ever-increasing demand to share information and integrate intelligence data. The explosive growth of information, specifically unstructured data, is making the detection and prevention of threats progressively more difficult. Consequently, the type of analysis needed is becoming more sophisticated, placing unique requirements on the tools used in these efforts (see Figure 1).

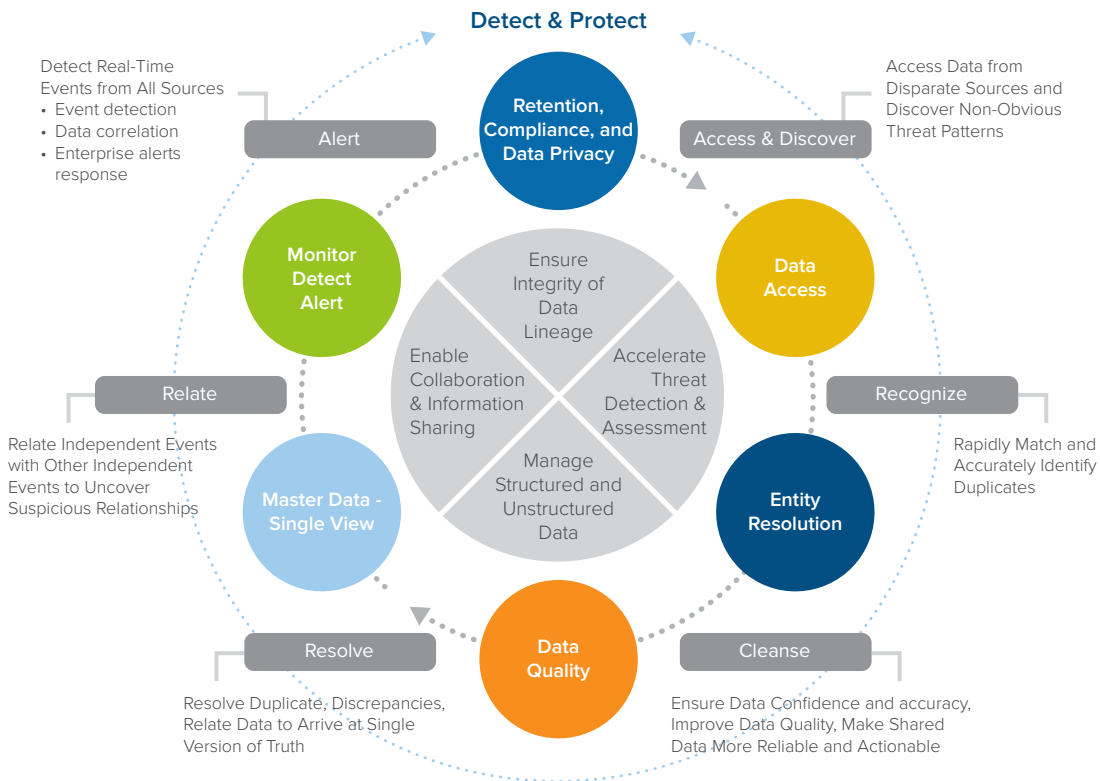


Figure 1: To transform data into actionable intelligence, U.S. intelligence agencies require software with sophisticated analytical capabilities to detect and protect against threats.

To transform large volumes of data into actionable intelligence, the intelligence community needs software that can:

- Detect suspicious activity by proactively identifying risks, anomalies, and unusual activities
- Deliver unprecedented insight through proactive monitoring and alerting
- Establish a data governance plan to use data efficiently across the organization

Let's take a closer look at each of these capabilities. Several capabilities are critical:

### Detecting Suspicious Activity by Proactively Identifying Risks, Anomalies, and Unusual Activities

- **Recognize: rapidly match and accurately identify duplicates.** Data formats, quality, validity, trustworthiness, and completeness vary greatly across systems. The ability to confirm identity is crucial to quickly detecting and resolving financial relevancy, unobvious affiliations, and reliable events to identities.
- **Resolve: merge duplicates to create a trusted, single version of the truth.** Golden records enable government organizations to seamlessly integrate, evaluate, and access data from multiple data sources to determine the most reliable, accurate, timely, and comprehensive master file across individuals, groups, events, and transactions.

### Delivering Unprecedented Insight Through Proactive Monitoring and Alerting

- **Relate: reveal relationships, patterns, associations, and relevant hierarchies.** Entity linking, relationship resolution, and identity matching enable agencies to sort and mine relevant data across disparate systems. The goal is to create a single, comprehensive view of people, events, organizations, and transactions.
- **Alert: correlate data that could indicate suspicious activities or threats.** The ability to apply user-defined analytical rules is essential to exposing hidden, distributed, and meaningful event patterns in unstructured or transaction data.

### Establishing a Data Governance Plan to Use Data Efficiently Across the Organization

- **Govern: maintain full and accurate records.** Agencies must keep records confidential, secure, and accurate. They must also carefully adhere to retention mandates and other compliance requirements.
- **Deliver: synchronize trusted master data to downstream applications and data warehouses.** By delivering accurate, timely, and aggregated data to analysts and data scientists, government organizations gain a comprehensive and trusted view.

## The Informatica Platform

Informatica has established itself as a leader in developing products that directly support missions in intelligence gathering. Specifically, we have extensive experience in:

- **Supporting intense analysis of people, organizations, and watch lists** by persistently searching, matching, deduplicating, and resolving data sets that contain information about high-risk individuals
- **Onboarding new, large, and complex raw data sets** in less than a day
- **Reducing organizational risk** by finding and analyzing data anomalies across data sources
- **Ensuring responsibility, transparency, and accountability** of data by capturing and tracking the pedigree and lineage of all data sources
- **Reducing annual data management costs** by 60 percent by eliminating the need for error-prone hand coding and consolidating multiple data sources into a single and common environment
- **Improving efficiencies of enterprise data management initiatives** through reuse of sophisticated data mappings and services
- **Autonomously collecting and correlating multiple data sources** on behalf of intelligence analysts
- **Broadening enterprise search and discovery capabilities** by providing JWICS and SIPR data integration and data quality services for TIRs, IIRs, and other complex unstructured data sources

To address the critical data analysis needs of the U.S. intelligence community, Informatica offers a comprehensive enterprise data integration platform. Our platform discovers, distills, and delivers the information necessary to meet requirements for effectiveness, transparency, and accountability. Siloed, disparate data is costly to maintain, difficult to manage, and substantially limited in value. But with the Informatica Platform, data located throughout the enterprise and cloud, whether structured, unstructured, or semistructured, is fully integrated.

### Unified

The Informatica Platform is built upon shared data services, metadata, and data rules. These shared capabilities reduce costs through reuse, consistency and transparency, and reducing the impact of continual change. Unified features of the Informatica Platform include:

- **A single set of role-based tools**, designed for each role but shared among IT and the organization, including a single Web-based configuration and administration console
- **Common capabilities delivered as reusable, shared services**, such as data access, data profiling, data cleansing, and data transformation
- **A common set of data rules and metadata** for lineage and reporting, shared among different users across the enterprise

## Open

Designed for maximum flexibility and return on data, the Informatica Platform is built for change and integration with all data, hardware, software, and technology standards (see Figure 2). Open features of the Informatica Platform include:

- **Universal data access** to quickly and reliably reach all data from all sources, regardless of type, structure, or location (e.g., on premise, with a third party, or in the cloud)
- **Interoperability** that supports a broad variety of hardware platforms, operating systems, databases, and networks
- **Open APIs and SDKs** that easily extend data integration capabilities and integrate them with other applications and technologies to speed development

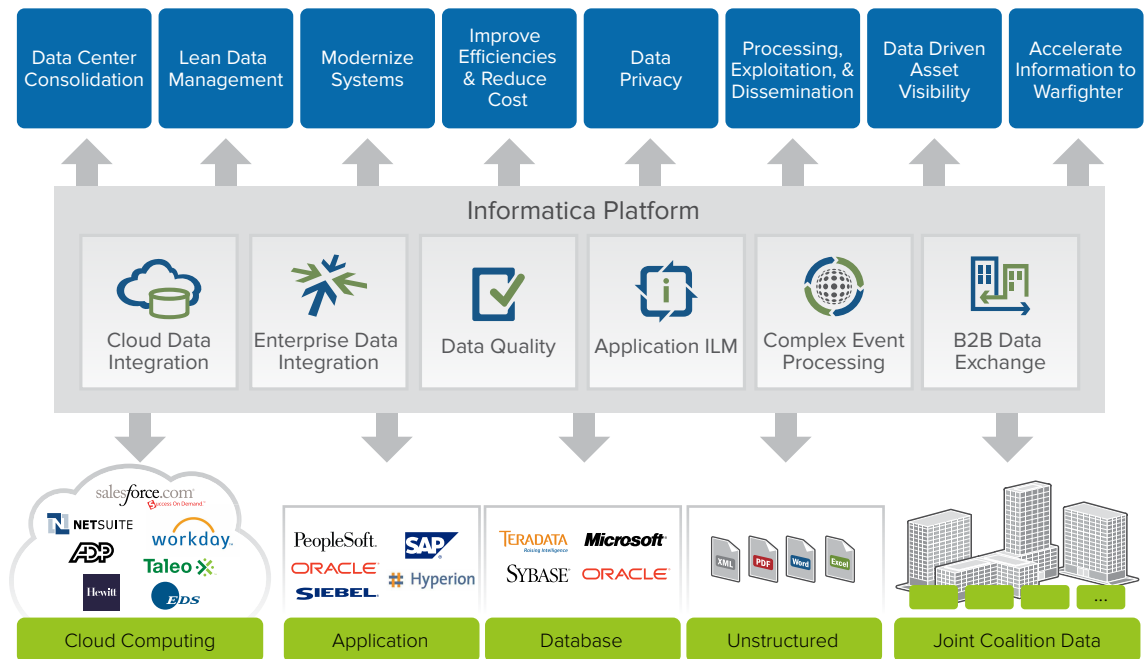


Figure 2: The Informatica Platform is built for change and integration with all data and all hardware, software, and technology standards.

## Economical

### Reduce Costs

- **Cut development and maintenance costs by 50 percent or more** by reusing existing Informatica skills and data integration logic for all data access across all projects and by combining data integration, federation, transformation, profiling, cleansing, and management on a single platform
- **Lower the cost of change** by insulating applications from underlying data changes



## Improve Efficiency Through Reuse

- **Reuse data integration logic and data services without any rework**, across all applications and for all types of data access—Web services, SQL, or batch
- **Seamlessly reuse data integration logic and data services** across a full range of data integration projects and use cases, including data warehouse augmentation, data warehouse consolidation, MDM hub extension, data migration, and SOA

## Boost Productivity at Lower Risk

- **Involve business users early** in data integration for faster project rollouts, early validation, and minimal rework
- **Save developers from redundant work across projects** by leveraging prebuilt components, a zero-code model-driven environment, and fully reusable data integration and data quality logic
- **Ensure compliance with data privacy policies** by relying on integrated data management capabilities that ensure that your data complies with all relevant policies and keep dirty data out of downstream applications and systems
- **Reduce outages, minimize costs, and meet SLAs** with such features as enterprise grade security for large, globally distributed environments, dynamic partitioning, and workflow on grid

## Comprehensive

### Mask Personally Identifiable Information

Informatica Dynamic Data Masking is high-performance, scalable data masking software. It enables intelligence organizations to block or mask sensitive or confidential information from unauthorized access to production or near-production databases, and seamlessly integrates with LDAP, Active Directory, and identity access management software. Key features include:

- Policy-driven real-time database security and monitoring
- Support for regulatory requirements such as
  - The amended National Security Act of 1947 and by the National Counterterrorism Center guidelines, as approved by the U.S. Attorney General
  - National Security Act of 1947, as amended
  - Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004, as amended
  - Homeland Security Act of 2002, as amended
  - Federal Agency Data Mining Reporting Act of 2007 (42 U.S.C. § 2000ee-3)
  - 18 U.S.C. § 2332b(f) (Acts of terrorism transcending national boundaries-investigative authority)
  - Executive Order 12333 of December 4, 1981, as amended, “United States Intelligence Activities”
  - Executive Order 13388 of October 25, 2005, “Further Strengthening the Sharing of Terrorism Information to Protect Americans”
  - Intelligence Community Directive (ICD) 501 of January 21, 2009, “Discovery and Dissemination or Retrieval of Information within the Intelligence Community”
- Versatile and unintrusive to applications or databases
- Integration with authentication software
- Real-time data masking and blocking

## Integrate Big Data

In simplest terms, the term “big data” means processes and procedures allowing an agency to create, manipulate, and manage very large data sets (e.g., terabytes or even petabytes) and storage facilities. The need to manage and make best use of that data is critical to the intelligence community in support of its gathering of information. Meeting the challenge of managing massive volumes of transaction data, multilanguage data sources, watch lists, and cross-domain GEOINT data is certainly daunting. When coupled with the explosive growth of new types of interactions, such as social media and real-time streaming data such as sensor technology, the task may seem insurmountable. Informatica seamlessly integrates these data sources and enables organizations like yours that gather intelligence to manage the complexity and volumes simultaneously.

Informatica PowerExchange® for Hadoop provides native, high-performance connectivity to the Hadoop Distributed File System (HDFS). It enables IT organizations to take advantage of Hadoop’s processing power using their existing IT infrastructure and resources.

## Enable Sophisticated Alerting

Informatica RulePoint® and Informatica Proactive Monitoring empower agencies to quickly set up, detect, and analyze the information they need to ensure the appropriate action is defined. Even if the data point clears this first test, examination of other transactions, events, and communications in conjunction with the original event can trigger alerts.

## Resolve and Match Identities

Informatica Identity Resolution is precise, high-volume software that searches identity data—and associated product, asset, and financial data—to identify matches accurately and quickly, regardless of language, structure, format, location, duplication, omissions, or errors. Using intelligent algorithms and strategies to locate identity data records inside large databases, Informatica Identity Resolution emulates an expert user’s ability to recognize matches. The powerful software also discovers connections among people, financial transactions, and passport information that might be hidden in the data to deliver reliably accurate search, match, or grouping results to your intelligence analysts.

## Retire Applications and Archive Data

Informatica Application Information Lifecycle Management (ILM) software enables your intelligence agency to cost-effectively handle data growth, safely retire legacy systems and applications, optimize test data management, and protect sensitive data. By handling every phase of the data lifecycle, your IT teams can improve information lifecycle management and better support application optimization, regulatory compliance, and IT consolidation. Key products in this family include:

- Informatica Data Archive
- Informatica Data Subset
- Informatica Data Warehouse Advisor™

## Consolidate Data Centers

Informatica ILM software supports your data center consolidation by intelligently identifying and relocating older and inactive data or legacy applications, while preserving the integrity of the data and application and data pedigree for future access. Informatica ILM software can reduce your intelligence organization's data center footprint, and eliminate your costs of maintaining backups and powering and cooling servers.

## Leverage Ultra Messaging

Informatica Ultra Messaging® delivers extremely low latency messaging over any network. As data volume continues to grow, the imperative to increase throughput requirements and also minimize latency is critical to support timely threat detection. Informatica Ultra Messaging provides a messaging bus to furnish the connectivity for all processes. The messaging bus employs a direct peer-to-peer, brokerless form of messaging whose architecture supplies great scalability, extremely low latency, and high throughput. The removal of a centralized broker component for messaging eliminates a key throughput bottleneck and latency overhead in contrast to legacy messaging architectures.

# The Informatica Platform in Action

The Informatica Platform has helped agencies in the intelligence community to surmount the challenges associated with gathering, categorization, handling, storage, and securing of data. It has provided a cost-effective, secure, and flexible platform for enabling their IT organizations to handle the most complex data requirements and to meet stringent retention and privacy mandates. The following case studies illustrate how the Informatica Platform enabled two agencies to accomplish their missions.

## Case Study 1: Onboard Data Faster

The agency established for itself the objective of onboarding a single data source in 90 days. But some challenges impeded this goal. One of these was the lack of any data quality protocol, which prevented the agency from being able to clean test data from its production database prior to onboarding the single data source. This in turn prevented the agency from complying with retention and deletion regulations. Another challenge was the organization's reliance on expensive error-prone hand coding initiatives, due to the lack of reusable mappings or services and the incidence of duplicate data sets.

The Informatica Platform enabled the agency to surmount these challenges. With it, the agency was able to:

- Onboard new, complex raw data sets in less than a day
- Pinpoint and prevent data quality issues by cleansing data during ingestion—before any problems could spread into production data warehouses
- Ensure the accountability of data by seamlessly integrating extract, transform, and load (ETL) and data quality services
- Implement a “create once, deploy everywhere” model for ETL and data quality mappings, services, rules, and logic
- Eliminate the need for hand coding by creating and populating codeless data structures and an unstructured data transformation service
- Project a 60 percent reduction in costs

Not only did the Informatica Platform accelerate the agency's delivery of data, but it also met the needs of the agency's intelligence analysts while reducing the costs associated with enterprise data management.

## Case Study 2: Optimize Enterprise Search & Discovery

This agency had several challenges hampering the normal course of its operation. The agency was ingesting up to 32,000 new unstructured files per day but had to make data available for search and discovery within 10 minutes of reaching its targets. It had unique unstructured sources such as TIRs, IIRs, and various message traffic feed, but no ability to merge, match, or deduplicate data sources.

Using a suite of products from the Informatica Platform, the agency met its challenges and was able to:

- Eliminate the need for hand coding with a codeless unstructured data transformation service
- Render all data discoverable by analysts within 10 minutes
- Seamlessly integrate ETL and data quality services
- Create “once”, deploy “everywhere” model mappings
- Merge, match, and deduplicate during ingest process
- Achieve superior data quality

The Informatica Platform succeeded in broadening the agency’s enterprise search and discovery capabilities. It also modernized the organization by providing a comprehensive JWICS and SIPR unstructured data integration and data quality service.

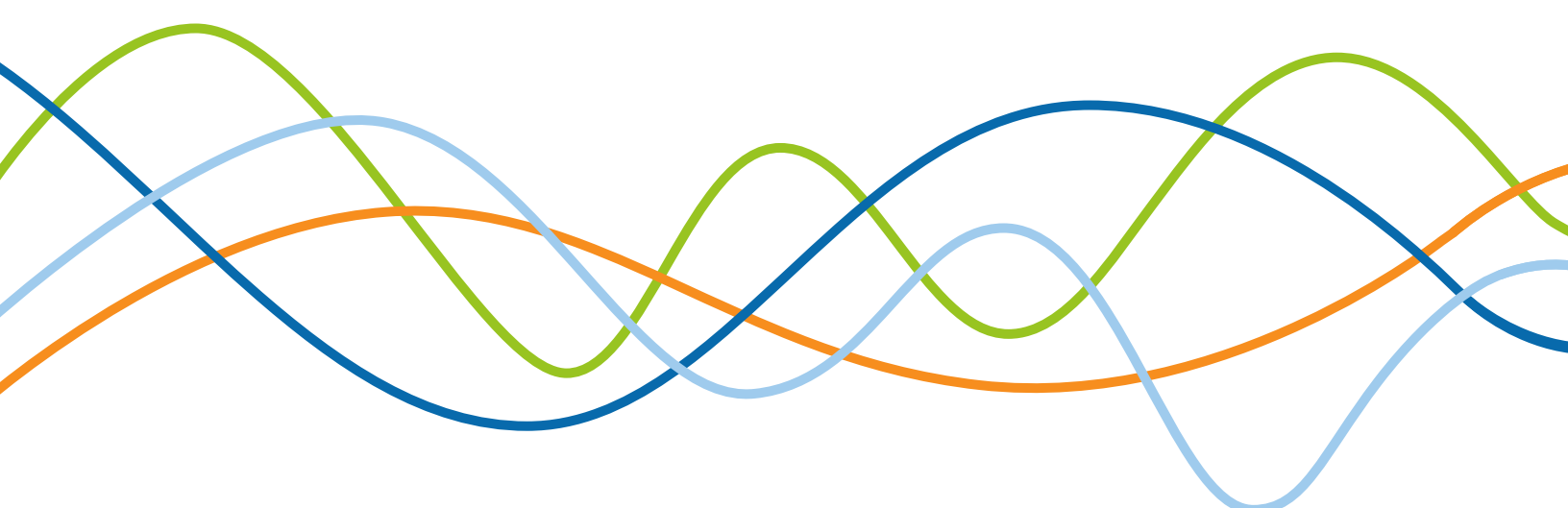
## Conclusion

Organizations in the U.S. intelligence community struggle to keep pace with the increasing deluge of data from multiple sources. Analysts must interpret data proactively and with unparalleled precision and sophistication. They require cost-effective, secure, and flexible software to handle the most complex data demands and to meet stringent retention and privacy mandates.

With its enterprise data integration platform, Informatica leads the industry in developing products that directly support the gathering of intelligence. The Informatica Platform discovers, distills, and delivers the information needed to meet requirements for effectiveness, transparency, and accountability. The intelligence community has used this open, unified, economical, and comprehensive data integration platform to streamline the chaos and complexity of data located throughout the enterprise and cloud, reducing costs and accelerating time to value.

## ABOUT INFORMATICA

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica for maximizing return on data to drive their top business imperatives. Worldwide, over 4,630 enterprises depend on Informatica to fully leverage their information assets residing on-premise, in the Cloud and across social networks.



**INFORMATICA**<sup>®</sup>

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  
[informatica.com](http://informatica.com) [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) [twitter.com/InformaticaCorp](https://twitter.com/InformaticaCorp)

© 2013 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and The Data Integration Company are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.