HRG
HARVARD RESEARCH GROUP
Experience - Expertise - Insight - Results

HRG Assessment:
Informatica's Intelligent Data Platform (IDP)

Founded in 1993 Informatica went public April 29, 1999 on the NASDAQ Stock Exchange. August 6, 2015, in what Harvard Research Group sees as an excellent strategic move toward long term growth, Informatica and their investors took the company private and INFA ceased trading on the NASDAQ. Informatica's product portfolio is focused on, data integration, ETL, information lifecycle management, B2B data exchange, cloud data integration, complex event processing, data masking, data quality, data replication, data virtualization, master data management, and ultra messaging. Informatica's customer base comprises more than 5,000 companies.

The Informatica Intelligent Data Platform (IDP) is the centerpiece of the company's reinvigorated strategic vision. IDP enables seamlessly integrated cross enterprise data management, bridging the gap between previous generation and current generation data management architectures. Focused around delivering the full potential of data fueled Cloud, Big Data, and IoT the Intelligent Data Platform offers a good way forward for customers looking toward the future.

The Informatica Intelligent Data Platform represents a rebranding of the Informatica Platform which has been in use since PowerCenter, Informatica's flagship product, was first released. IDP is represented in code however IDP is not a single entity. IDP is a combination of metadata (Live Data Map), task specific engines, and optimizers the combination of which defines IDP. Live Data Map is used extensively in Secure@Source and Intelligent Data Lake. IDP also supports streaming data-based solutions.

Core products and services that leverage IDP capabilities include Secure@Source, Informatica Cloud, Informatica Big Data, Data Quality, Data Integration, Master Data Management (MDM), MDM 360, Informatica Data Lake, and the Informatica Enterprise Information Catalog (EIC). Informatica is a data management company focused on integrating and conditioning data. Informatica moves data but does not store or contain data with the exception of Informatica's MDM. IDP and the products and services built to work with IDP make data available for use with a wide range of applications and platforms. IDP, an intelligent data aware platform, uses the power of metadata to streamline business transformation and make Data and IT resource integration efficient and cost effective.

IDP platform services-include metadata management services, Live Data Map (index of external, enterprise-wide metadata), security, scheduling, code management, and in-box processing engines. IDP presents a common metadata layer for integrating Informatica solutions, a user experience for different user persona, common engines across Data Integration, Data Quality, Virtualization, B2B and other solutions. There is a common workflow engine across the platform. IDP data processing engines include Informatica's own DI, DQ, and Parsing engines as well as code generators for SQL, HiveQL (Map-Reduce, Tez), Spark, and Spark Streaming.

Data Integration

The current generation of data management systems must be always on, always available, easy to customize, easy to use, and platform agnostic. Data is the fuel that businesses run on and data must be high quality and continuously available. Data management solutions must be able to handle structured, unstructured, and poorly structured data while continually collecting, cleaning and refreshing data stores. Solutions that can organize, sanitize, and make sense of the increasing flows of data are essential if we are to realize the full benefits of Cloud Computing, Big Data and the Internet of Everything.

Business Transformation

Business transformation requires cross-enterprise data integration. Beating the competition is key; getting more out of your data in order to advance the business is a requirement. Effective data integration drives real long term cost reductions, improved operational efficiency, enhanced competitive advantage, streamlined regulatory compliance, and increased productivity. Today data in corporations comes in an increasing variety of formats, types, and forms including structured and unstructured data harvested from diverse heterogeneous environments.

Data is no longer simply structured data artifacts sitting in a data mart or data warehouse to be used in generating a constant stream of reports or continuously answering a preconfigured set of executive queries. Metadata helps identify and find what data is available and useful by leveraging machine based inference / learning and pattern matching to analyze and correlate all available metadata. Great quality continuously available data is the goal resulting in more effective use of people, physical infrastructure, energy, and a better ROI for valuable data resources.

Until recently data has been collected and analyzed after the transaction or engagement has completed. This data is historic, static, and an archived artifact of a past event. Today Big Data, Social, and sentiment analysis are driving customers to transform their use of data to include archived structured data and real time or live structured and unstructured data. We are now seeing the transition from the use of data to reduce cost, reduce risk, increase profit and increase productivity to the use of real-time and historical data to influence the outcomes of a wide range of transactional engagements between customers and businesses. Data that is useful while an engagement is underway is powerful as it facilitates upselling through the introduction during the engagement of add-on offers with a high degree of relevance to the engagement at hand ensuring customer satisfaction and loyalty. Much of the data fueling and enhancing the customer experience on mobile and web based devices is unstructured and poorly structured data from sources like Facebook, Twitter, LinkedIn, Pinterest and Instagram.

Time to Result

Time to result can determine the success or failure of an endeavor. Now when an application does not respond or deliver results in near real time customers are likely to switch to another provider. The current generation of web based applications learn from and influence engagements in real time as transactions are occurring. Web and Mobile based applications increasingly use historical customer data combined with real time engagement data to deliver a range of up-sell opportunities to customers while transaction are occurring.

Real Time Customization

The real-time customization of offers to customers at the point of engagement on line or on the phone when placing an order or calling in to make changes in their billing information (for example: make an address change, or change their credit information) is powerful. New and evolving customer experience applications require the near real-time integration of distributed heterogeneous data sources. Achieving and maintaining a clear competitive

advantage drives the increased requirements for cross enterprise data integration. Beating the competition is key and getting more out of your data in order to advance the business is a requirement. Additional considerations include: data divergence resulting from increasing globalization, increasing competition from all corners of the globe, government focused compliance requirements, risk mitigation, and the elimination of operational risks (lowering the risk of being in business).

In the past Business Intelligence based decision-making was based on static historical data. Today Operational Analysis and decision-making are increasingly based on high velocity, high volatility, real-time data collected in a wide range of formats from all across the enterprise. This data has to be extracted, transformed, cleaned and made accessible in a relational format, a files system format, or through a web services format depending on specific application requirements.

The Informatica Intelligent Data Platform eliminates recoding and ensures that no quality problems are introduced as part of the data integration process. Informatica addresses data profiling and quality issues, such as parsing, address standardization, and matching. Enterprises will benefit from adopting a metadata focused approach to business and data transformation. The Intelligent Data Platform takes advantage of Informatica's unique insight into metadata to unlock data value without having to move the data. The Informatica Data Integration Platform leverages metadata across business intelligence and data modeling tools, packaged applications, and messaging technologies, increasing efficiency and improving accuracy throughout the enterprise. Informatica products and technologies share a common set of metadata through a centralized metadata catalog, the Informatica Enterprise Information Catalog (EIC).

PowerCenter - Metadata Manager

In 2016 data ecosystems are neither static nor contained. Metadata (data about data) captures important information about the enterprise data ecosystem. Metadata describes the structure and workings of an organization's use of information and the systems it uses to manage that information. The modeling techniques that produce and manage metadata and provide a complete description of the systems within an enterprise are data models, process models, application models, and requirements.

Informatica's PowerCenter Advanced Edition provides access to the full suite of products that comprise the Intelligent Data Platform offering. Informatica's metadata driven approach to the identification and integration of cross-enterprise data reduces TCO. Minimal, if any, recoding is required as the metadata for the source and target structures is already in the system and can be reused. A significant portion of the implementation details can be automatically generated. The simplification and automation the Meta Data Manager enables is readily apparent when managing complex data environments.

Metadata Manager aggregates and links metadata from a mix of heterogeneous data sources providing context to IT artifacts, insight into data relationships, and visibility into the potential impact of data changes, thereby reducing delivery time and helping IT teams obtain more accurate cost estimates through its data lineage and impact analysis capabilities. Customers can use Metadata Manager to browse and analyze metadata from disparate metadata repositories to understand how information and processes are derived, and manage the relationships between them.

Enterprises will benefit when they adopt a metadata approach to defining and modeling data driven business transformation. Business users can now have a single view of customers and products across the organization. Additionally, they will be able to pull together metadata and data (using Vibe) from multiple different sources in order to meet compliance requirements without having to create more work / cost by pushing compliance down to individual divisions of the company.

Enterprise Information Catalog

The creation of the Enterprise Information Catalog (EIC) populated with metadata models for the analysis and design that drive data integration is central to enabling the full capabilities of the Intelligent Data Platform. The metadata layer (catalog) is populated with the metadata models used to create the analysis and design that drive data integration.

Vibe

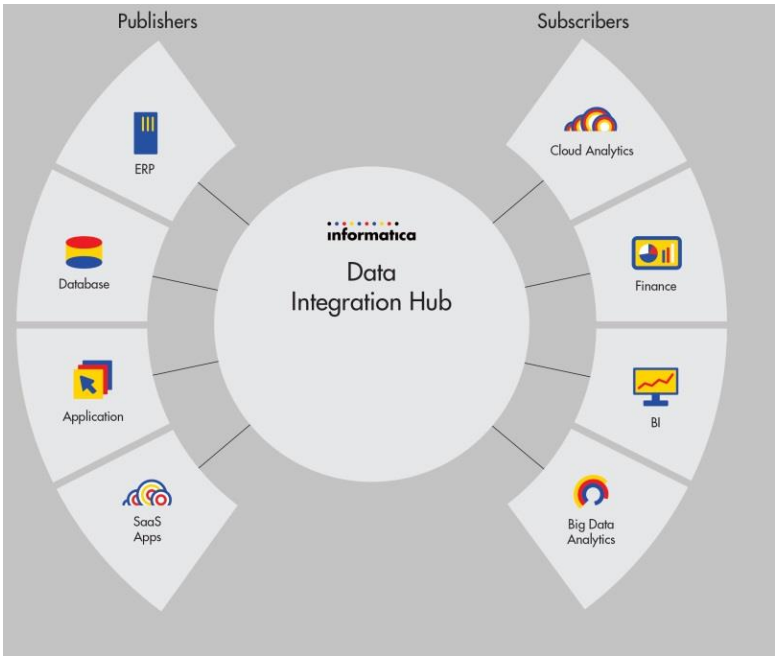
The Informatica Intelligent Data Platform, powered by Vibe, supports reuse of data integration mappings created for users or departments across cloud environments, enterprise data centers, and big data platforms. Vibe enables access to data, regardless of location, format or origin. Vibe is an engine not a product but a capability and a core part of IDP that Informatica utilized in designing and launching PowerCenter their flagship product.

Vibe, the core component of Power Center, separates business logic from operations execution. Organizations can run vibe enabled applications and apps on any platform, anytime, and anyplace without having to modify business logic. Vibe, a virtual data machine, separates and abstracts the logical data design or schema from the platform centric execution instructions related to doing ETL and other integration tasks. This separation of design logic from execution platform specifics shields customers from unnecessary complexity freeing them up to allocate scarce highly skilled human resources to more critical tasks.

Vibe enables the separation and abstraction of logical design elements such as a data mapping or a data quality rule from platform specific execution code. Vibe provides the functionality that enables use on a variety of data platforms such as Oracle or Hadoop without requiring a unique design effort for each unique platform. Vibe separates the logical design from the design execution platform and it is this aspect of Vibe that shields customers from the majority of underlying data platform specific technical complexity.

Informatica Data Integration Hub

Informatica's Data Integration Hub (DIH), uses the publish/subscribe Hub and Spoke model for data integration where the spokes are either publishers, subscribers, or both facilitating data integration automation and maintaining strict control. This approach lets customers decouple data sources and destinations and leverage logical models allowing application owners to self-subscribe to a centralized catalogue of published data sets.



Through centralized data management and available Informatica Data Quality, all data can be certified and cleansed prior to publication to eliminate the risk of bad data proliferating across applications, departments, cloud applications, or analytical systems.

DIH's wizard based interface and task centric dashboards for operations and IT facilitate agility while data lineage based impact analysis enables future focused decision making for data management and

integration environments. This GUI approach to data management enables less technical users to intuitively manage connected applications, publications, and subscriptions. Catalogs of available publications and auto-mapping enable self-service for rapid on-boarding for new applications.

Data Integration Hub's topic specific persistent storage is managed through configurable retention policies. Published data is maintained either until all the consuming applications has received the data, until its retention period has expired, or for long-term storage archived in a data lake. With DIH applications publish their data once resulting in reduced overhead/, DIH constrains costs associated with SaaS apps that charge on a per API call basis every time they are synced. Downstream data consumers benefit directly as DIH decouples redundant point to point interfaces and corrects for logic duplication and inconsistencies. Legacy business applications, analytics systems and data stores that have been hard-wired for point-to-point connectivity may cause management and maintenance problems because of the confusing mix data flows and redundant connections. Consumers of downstream data will benefit as DIH's publishes data sets that have been validated and enrichment operations are executed only once to ensure that downstream data consumers and systems receive only certified consistent data. New projects can now subscribe to existing data domain definitions and publications using DIH and inherit the full benefit of standardization and consistency checking implemented on initial publication.

Informatica's Data Integration Hub assists non-technical users in generating data flow integrations that would previously require a developer. DIH functions as an orchestration engine supporting hybrid environments and invoking publication and / or subscription logic for execution. DIH integration with Informatica Cloud and Big Data Management automates data management and integration across cloud, big data and traditional systems moving data between cloud and on premise systems allowing publication and subscription workflows to run on the platform. DIH ensures enhanced organizational agility and better data governance.

Secure@Source

Secure@Source works with other installed security technologies to help organizations identify and locate sensitive and confidential data which is critical for highly regulated industries and government agencies. Secure@Source only allows the movement of secure data known secure platforms using predefined rules designed to meet regulatory compliance and information security requirements. Company access policies are linked to compliance, regulation and governance requirements. Users can align security rules like data masking, alerting, blocking, encryption and tokenization to data sets. Customers can determine which data stores need to be more secure in order to manage risk and exposure.

Informatica Secure@Source identifies, analyzes and visualizes data relationships, proliferation and sensitivity. Secure@Source secures data at the source or origin, before it is copied or distributed, throughout its lifecycle, and analyzes and visualizes data for risk prioritization using a data risk heat map. Customers monitor data risks and vulnerabilities to prevent breaches and insider abuse. The solution shows what data has value, where the data resides and how it flows across the organization. Using Informatica's data discovery, profiling, analysis, and visualization technologies customers can monitor and control sensitive data.

When using IDP with Hadoop organizations can identify the potential for risk or fraud by correlating, and analyzing cross enterprise data. With IDP and Informatica Secure@Source financial institutions can now observe their customers ATM, POS, and on-line use patterns to identify anomalies for indications of bank card and identity theft.

Intelligent Data Lake

Customers need a mechanism like Informatica's Intelligent Data Lake (IDL) to collect, discover, refine, govern, provision, securely share, and control all data assets and internal and external feeds. The Data Lake provides fast

and easy access to clean data for analysis and helps facilitate data management, governance and policy compliance for IT departments. IDL speeds up data provisioning and enables automated cataloging, as well as, low-latency and scalable storage and processing.

Master Data Management (MDM)

Informatica Master Data Management (MDM) establishes a master data file linking enterprise data assets and helps control and coordinate data records across a heterogeneous mix of architectures, platforms, and applications. Benefits include long term cost reductions, improved operational efficiency, effective competitive positioning, efficient regulatory compliance, and enhanced productivity.

IT and business professionals are increasingly focused on the integration of complex, highly distributed and heterogeneous cross-enterprise data flows into a single MDM view. Cross-enterprise data integration and digital business transformation can deliver substantial business value. Data integration drives and enables improved operational efficiency, enhanced competitive advantage, streamlined regulatory compliance, and increased productivity. Stand-alone data integration solutions that have worked in the past are not the best choice for today's enterprise-level data integration requirements.

Machine Learning

IDP uses both supervised and unsupervised machine learning. With supervised learning techniques for parsing semi-structured files users train the parsers with sample files and the engines automatically handle files of varying formats. With the unsupervised learning clustering technique fields, columns, and objects which represent similar concepts are grouped together. Recommendation engines (for suggestions to users) are another use of machine learning techniques. Core techniques are implemented as modules which are then extended based on a specific use case.

Conclusion and Recommendation

Cross-enterprise data integration and digital business transformation deliver substantial business value. Business transformation requires cross-enterprise data integration and cross enterprise data management to bridge the gap between previous generation and current generation data management architectures. Informatica's Intelligent Data Platform (IDP) is an essential technological innovation focused on delivering the full potential of Cloud, Big Data, and IoT. IDP uses metadata from business intelligence and data modeling tools, packaged applications, and messaging technologies to enable business transformation and facilitate Data and IT resource integration. Informatica's Data Integration Hub (DIH) uses metadata to define and model data driven business transformations. The DIH's wizard based interface and task centric dashboards provide business users with a single view of customers and products from across the organization and enables less technical users to manage connected applications, publications, and subscriptions.

As organizations flatten and simplify their internal IT infrastructure using converged and hyper converged systems for energy efficiency many of these organizations are shifting to a hybrid cloud based IT infrastructure to increase productivity and reduce cost. Those organizations that are moving to the Cloud will gain significant advantage by adopting Informatica's Intelligent Data Platform and associated products to help position and transform themselves to take full advantage of Cloud, Big Data, IoT. If you are not currently considering Informatica as a strategic solution partner HRG strongly recommends that you perform your own evaluation – you will be glad you did.

Harvard Research Group is an information technology market research and consulting company. The company provides highly focused market research and consulting services to vendors and users of computer hardware, software, and services. For more information please contact Harvard Research Group:

Harvard Research Group
Harvard, MA 01451 USA

Tel. (978) 456-3939

e-mail: hrg@hrgresearch.com

http://www.hrgresearch.com

