Data Sheet

Informatica Data Integration Hub

Benefits

- Modern hub-based hybrid architecture to connect big data, cloud and traditional systems
- Centralized, automated data management for Hadoop data lakes and hubs
- Hybrid publish and subscribe data integration for SaaS and on-premise applications
- Connectivity to all PowerCenter and Informatica Cloud sources and targets
- Wizards and web-based user interface for self-service and enhanced productivity

The world's first modern hybrid hub-based data integration to connect big data, cloud and traditional systems

The explosion of big data, SaaS (Software as a Service) and on-premise enterprise applications and analytical systems creates complexities, challenges and fragmentation for companies needing to rapidly evolve their systems to deliver competitive advantage. Organizations struggle to integrate their big data, cloud and on-premise systems to deliver fresh, clean data for analytics and operational applications while managing hundreds of interdependent point-to-point integrations. With organizations increasingly adopting SaaS apps, companies need ways to manage hybrid data flows to link on-prem systems with the cloud. CIOs and IT teams across the globe want to take advantage of more efficient new technologies, new on-premise and cloud-based enterprise applications and the opportunities presented by data lakes.

Modern architecture for managed big data – cloud integration

Informatica Data Integration Hub empowers large organizations to embrace change and the opportunities of new applications and analytics systems while managing storage in Hadoop. The centralized modern hub-based architecture is the foundation for agile and managed enterprise data integration. As the first to apply a publish/subscribe model to data integration linking big data, cloud and traditional systems, Informatica delivers productivity and intelligent automation without compromising control.

The Data Integration Hub simplifies the delivery of fresh data to all analytics systems and application-to-application data integration so that organizations can support any volume, format, latency, or protocol within a single data integration platform. And because it's a hub, data management, monitoring and control are all centralized in a web-based console. This ensures that data moving through the hub is trustworthy, secure, and traceable.

Orchestrating Informatica's on-premise, big data and cloud connectivity, data transformation and quality with a central pub/sub architecture enables companies to put visionary hybrid solutions into production. Data Integration Hub builds on the rich, high-performance on-premise big data and traditional connectivity of PowerCenter with SaaS connectivity of Informatica Cloud in one hub-based architecture. For example, an organization could source data from Salesforce, on-premise applications, databases and mainframes, process the data, automatically store it on their Hadoop data lake and then load curated data onto all of their analytics and operational systems, including Amazon Redshift and Teradata.

Wizards, web-based user interfaces and the ability to leverage PowerCenter and Informatica Cloud mappings results in new levels of team collaboration and productivity. These make the universal connectivity, rich transformation capabilities, scalability and data quality that Informatica is famous for easier to use and accessible to more team members.

Data Integration Hub

The centralized approach with easy to use interfaces and visibility across data flows is even more valuable for notoriously complex big data. Big Data Integration Hub version also includes integrated support for Hadoop repositories, such as Cloudera and Hortonworks. Big Data Integration Hub abstracts the complexity of storing and managing raw and curated data sets in a Hadoop data lake or hub. Data Integration Hub indexes all of the data it stores in Hadoop with Hive to make the data accessible and queryable by analytics tools and other applications.

Publish-and-Subscribe puts an End to Point-to-Point Complexity

Using this innovative approach, data sources are decoupled from destinations, enabling applications to publish once but effortlessly support one-to-many consuming applications and analytics systems. The Informatica Data Integration Hub addresses the lack of management and data linage over traditional point-to-point integrations that, over time, can drive up the total cost of ownership and compromise the ability of IT to serve the business. The Data Integration Hub changes all this by empowering you to determine which type of integration pattern and which type of publication repository (RDBMS, Hadoop, file) best suits the need for the lowest cost with full data transparency end-to-end.

Deliver New Integrations Faster with Less IT Overhead

To reduce development time, the Data Integration Hub enables self-service with a web-based user interface to enhance productivity. Designed for the non-technical, a wizard guides users through steps to publish data to a central catalog or subscribe to the data in it. And because the data is managed centrally, it becomes easy to promote its reuse across different applications and analytics systems. With just a few clicks, users can combine, filter, and transform data to meet their specific needs.

Key Features

Self-Managed Big Data Management

Flexible Hadoop, file system and RDBMS persistency layer options enable published data to be retained until either all the consuming applications have received it, its retention period has expired or for long-term storage in a data lake. Publishing applications publish their data once; the hub then delivers that data to any number of consuming applications. This significantly reduces transactional system overhead and virtually eliminates process dependencies, enabling downstream applications to readily meet the needs of business users.

Persistence is self-managed by the Data Integration Hub. Persistent storage for each topic delivered by Data Integration Hub is automatically created and managed, making it effortless to store data and archive consumed data. In Hadoop, publication data is converted into Parquet format by Spark and indexed in Hive to be available to any analytics systems querying the data lake as well as applications that may want to access the data. Because all data in motion is persistent, interactions between applications are managed centrally, data management can be uniformly applied, and data integration processes can be monitored and controlled through a single user interface that abstracts sources and targets.

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

Hybrid Support for SaaS Applications

Modern publish and subscribe data integration for a hybrid world that can seamlessly move data between hundreds of supported cloud and on-premise systems. In addition to leveraging robust PowerCenter connectivity and processing Data Integration Hub also integrates with Informatica Cloud's broad array of SaaS sources and targets, including Salesforce, AWS services, Azure, Netsuite, Workday and many others. This flexible hybrid architecture accelerates data agility to unprecedented new levels of productivity and manageability.

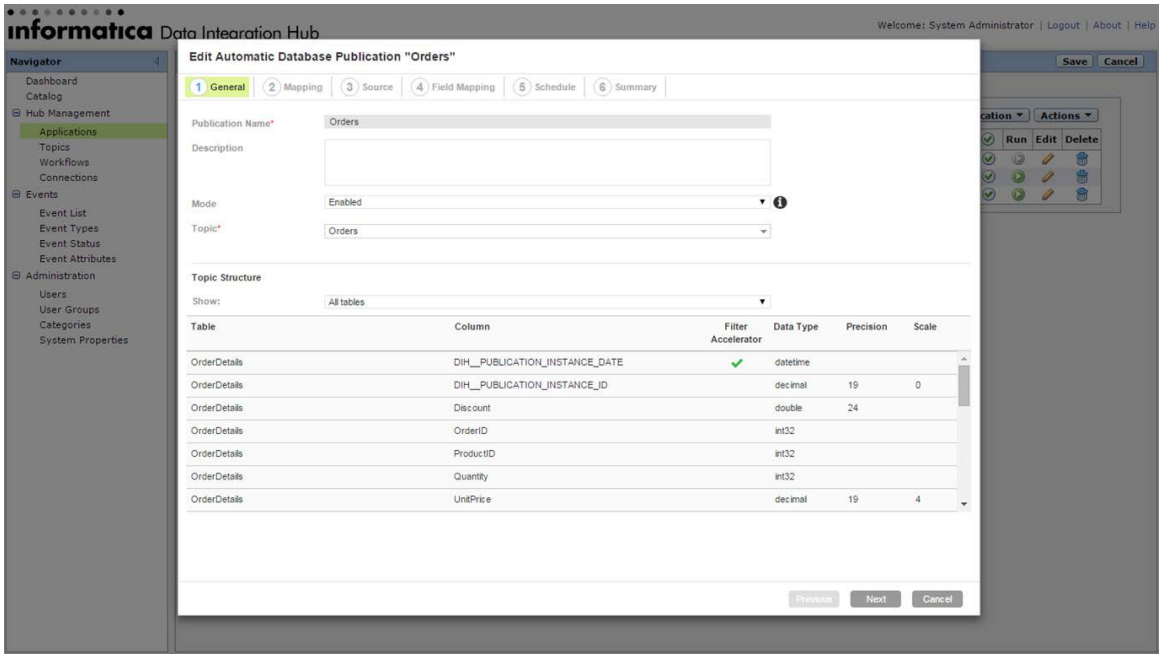
End-to-end centralized data flow management

Instead of IT building hundreds or even thousands of point-to-point integrations, publication can be once for any number of consuming systems. The interfaces between publishing and subscribing applications are now simply created and managed through interaction management. By decoupling data sources and destinations, applications are less process-interdependent, enabling downstream applications to better meet the data delivery needs of business users.

Through an intuitive subscription wizard, less technical users can use self-service provisioning capabilities to subscribe to the published data sets they are authorized to access. This greatly accelerates time-to-market for new integrations because it requires less IT involvement. This powers self-service and reuse for continuously evolving enterprise systems. Centralized management of data flows enables higher team productivity.

"Informatica is now productizing the Hub-and-spoke data integration pattern in their Data Integration Hub (DIH). The DIH provides the ability for multiple integration flows to re-use canonical data in a publish/subscribe paradigm and remove the point-to-point nature of traditional data integration."

—Stewart Bond, Senior Consulting Analyst, Info-Tech Research Group



Through a wizard, less technical users can manage connected applications, publications, and subscriptions. Catalogs of available publications and auto-mapping enable self-service on-boarding for new applications.

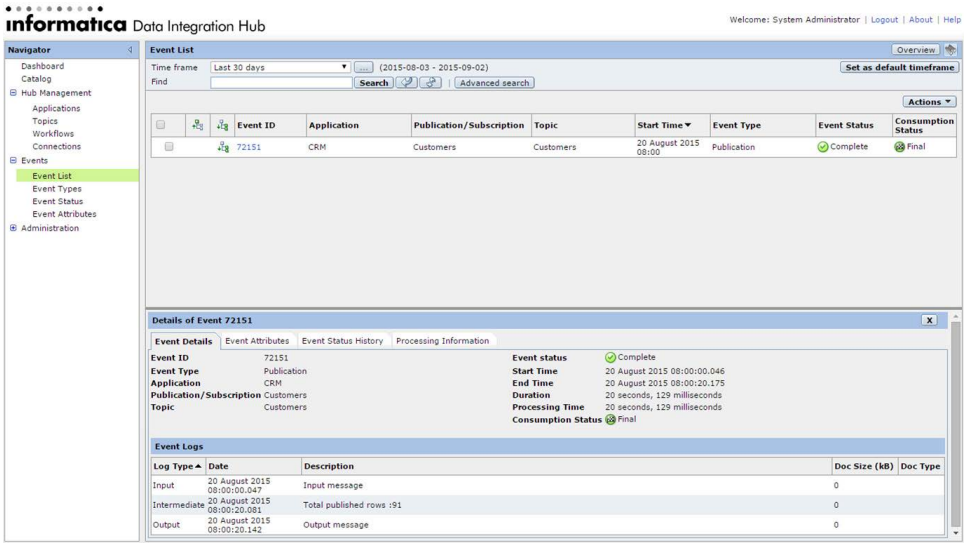
About Informatica

Informatica is a leading independent software provider focused on delivering transformative innovation for the future of all things data. Organizations around the world rely on Informatica to realize their information potential and drive top business imperatives. More than 5,800 enterprises depend on Informatica to fully leverage their information assets residing on-premise, in the Cloud and on the internet, including social networks.

Monitoring and Control

As data moves through the Data Integration Hub, events are captured along with associated metadata in an operational data store (ODS). Through a business-friendly UI, users can drill down into the details of any interaction, publication or event to see details and status. With role-based access controls for security, data access is based on centralized authorization.

Robust, configurable notification and alerting gives operators and application managers the peace of mind that they'll know of a problem before the business does. With configurable dashboards, operators can monitor and measure integration performance against service level agreements (SLAs) to ensure they meet business needs for data delivery. Better monitoring and centralized control can transform disconnected and independently created data flows into an orchestrated system to empower rapid change as systems are updated and modernized.



With built-in visibility, users can monitor and track status and receive alerts if there is a problem. Drill-down capabilities give access to details on every publication.

Data Integration Hub

informatica
Put potential to work.™

Worldwide Headquarters, 2100 Seaport Blvd, Redwood City, CA 94063, USA Phone: 650.385.5000 Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871 informatica.com linkedin.com/company/informatica twitter.com/Informatica

IN08_0915_02473