

# Informatica Data Integration Hub

## Accelerate Time to Value and Agility With the World's First Modern Hybrid Publish/Subscribe Data Hub

### Benefits

- Modern hub-based hybrid solution to connect and process data in cloud, big data, and traditional systems
- Visibility, monitoring, and alerting across workflows
- Publish/subscribe data integration automation for SaaS and on-premises applications
- Orchestration of data processing using PowerCenter, Informatica Cloud, or Big Data Management
- Wizards and web-based user interface for self-service and enhanced productivity

The explosion of big data, software as a service (SaaS), and on-premises enterprise applications and analytical systems creates complexities, challenges, and fragmentation for companies needing to rapidly modernize their systems to deliver competitive advantage. Organizations struggle to integrate their big data, cloud, and on-premises systems to deliver fresh, clean data for analytics and operational applications, while managing hundreds of interdependent point-to-point integrations.

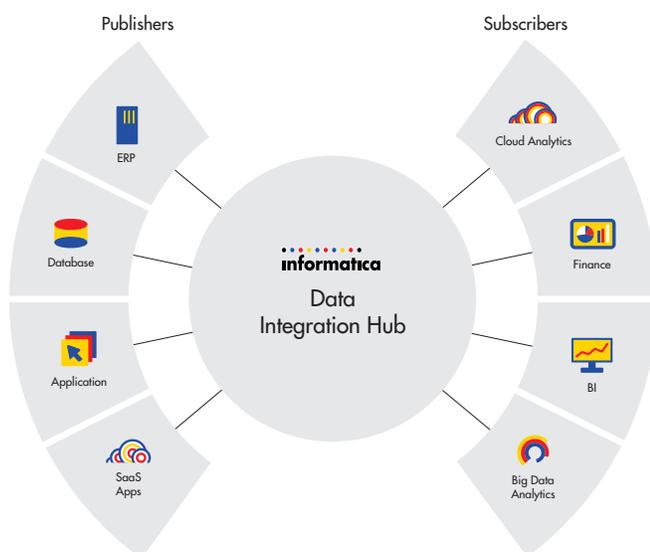
With organizations increasingly adopting SaaS applications, companies need ways to manage hybrid data flows to link on-premises systems with the cloud. CIOs and IT teams across the globe want to take advantage of more efficient new technologies like Hadoop and the advanced analytics opportunities presented 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 as well as relational database and file store options. 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 and clean 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, it centralizes data management, monitoring, and control in a web-based console. This ensures that data moving through the hub is trustworthy, secure, and traceable.

Organizations can put consolidated hybrid solutions into production by orchestrating Informatica's on-premises, cloud, and big data connectivity; data transformation; and quality with a central publish/subscribe architecture. Data Integration Hub builds on the rich, high-performance on-premises big data and traditional connectivity of PowerCenter with SaaS connectivity of Informatica Cloud in one hub-based architecture. For example, an organization could first source data from Salesforce, on-premises applications, databases, and mainframes, then process the data and automatically store it in their Hadoop data lake, and finally, deliver the curated data into all analytics and operational systems, including Amazon Redshift and Teradata.



Wizards; web-based user interfaces; and the ability to leverage PowerCenter, Informatica Cloud, and Big Data Management mappings result in new levels of team collaboration, productivity, and flexibility. 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 to facilitate business self-service, which is a chief goal of data-driven organizations.

This centralized hub approach combining easy-to-use interfaces and visibility across data flows is even more valuable for notoriously complex big data. Big Data Integration Hub Edition also includes integrated support for processing data at Hadoop scale and efficiency. Big Data Integration Hub abstracts the complexity of storing and managing raw and curated data sets in a Hadoop data lake or hub by indexing all of the data it stores in Hadoop with Hive to make the data accessible and query-able by big data analytics tools and other applications.

### **Publish/subscribe data hub a modern method to better manage complexity**

Using this innovative approach, data sources are decoupled from destinations, enabling applications to publish once and automatically format the data for each consuming system and deliver it at just the time it needs it. The Informatica Data Integration Hub addresses the lack of management and visibility that, in traditional point-to-point integrations, can over time drive up the total cost of ownership and compromise the ability of IT to serve the business at the speed it needs. The Data Integration Hub changes all this by empowering you to easily configure orchestrated data processing based on your choice of workflow engine and publication repository for persistence (RDBMS, Hadoop, file store) for greatest efficiency combined with full data transparency end-to-end.

### **Deliver new integrations faster with greater self-service**

To reduce development time, the Data Integration Hub enables self-service with a web-based user interface to enhance productivity. Designed to empower less technical users' participation, 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**

### **Hybrid support for SaaS applications**

Data Integration Hub provides modern publish/subscribe data integration for a hybrid world that can seamlessly move data between hundreds of supported cloud and on-premises 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, Amazon Web Services, Microsoft Azure, NetSuite, Workday, and many others. This flexible hybrid architecture accelerates data agility to unprecedented new levels of productivity and manageability.

### **Integrated Big Data Management**

The Big Data Integration Hub package empowers organizations to put Hadoop to work together with their traditional systems with automated data flows among all of them as well as the cloud. Data Integration Hub can fully leverage all of the high-performance processing options on Hadoop of Informatica Big Data Management: Blaze, Spark, Tez, and Map/Reduce. With Data Integration Hub, you can configure publication and subscription workflows to execute on the platform and engine that is best suited to the use case and goals.

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. Flexible Hadoop, file system, and RDBMS persistency layer options enable published data to be retained either until all the consuming applications have received it, until its retention period has expired, or for long-term storage in a data lake. Publishing applications publish their data once and after executing

transformation and validation logic once, the hub then delivers that data to any number of consuming applications for greater efficiency and consistency. This significantly reduces transactional system overhead and virtually eliminates process dependencies, enabling downstream applications to readily meet the needs of business users.

In Hadoop, publication data is converted into Parquet format by Spark and indexed in Hive to be available to big data analytics and applications. 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.

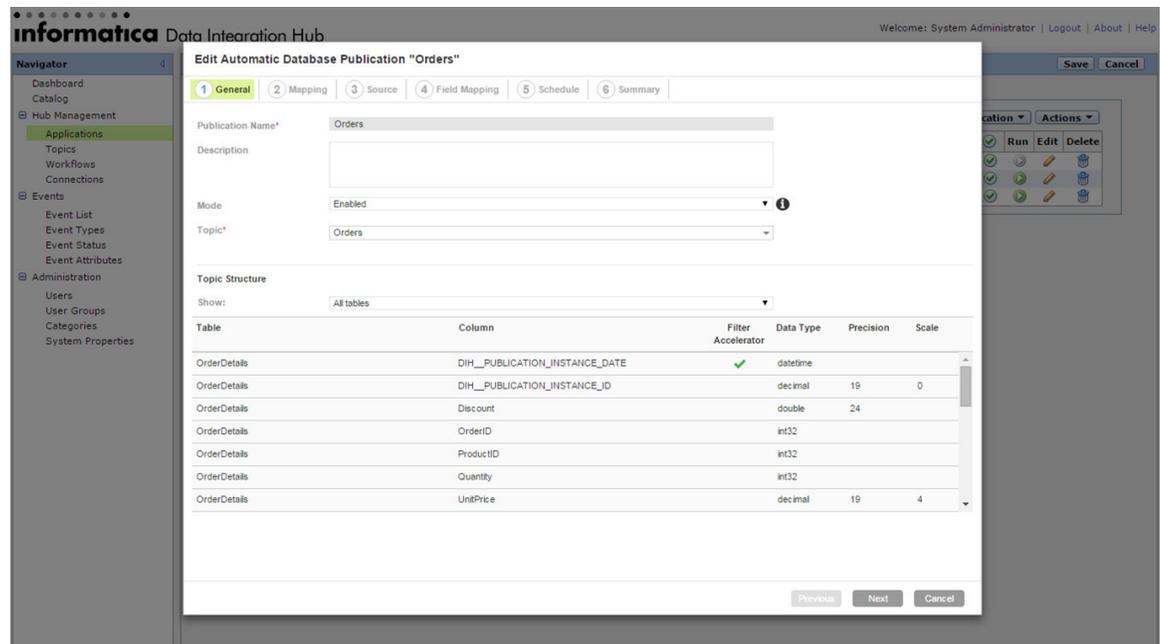
Through centralized data management and available Informatica Data Quality, all data can be certified and cleansed prior to publication. This virtually eliminates the risk of bad data proliferating across applications, departments, cloud applications, or analytical systems.

### End-to-end centralized data flow management

Instead of IT building hundreds or even thousands of point-to-point integrations, data is processed and published once, then delivered to all subscribing systems at the time and in the formats each one needs. 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.

“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

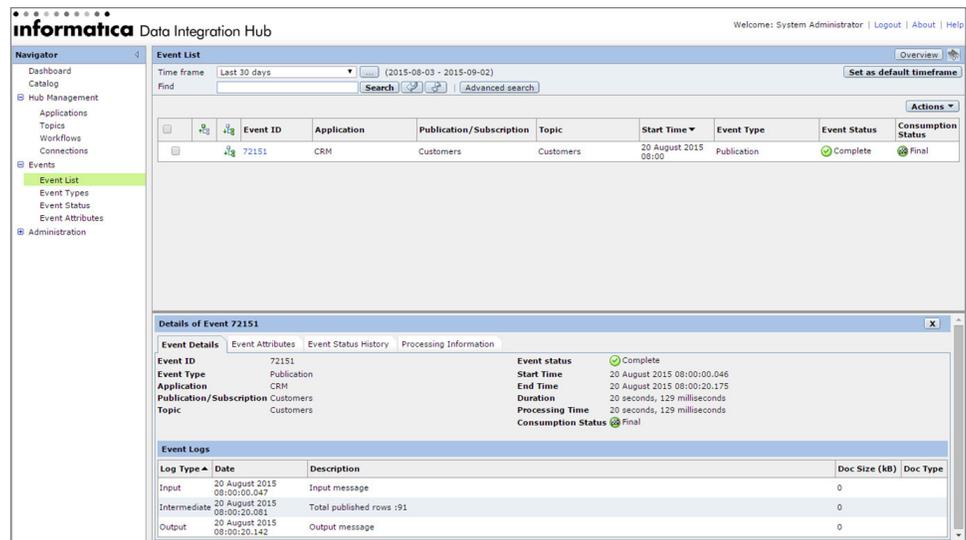
Informatica is a leading independent software provider focused on delivering transformative innovation for the future of all things data. Organizations around the world rely on Informatica to realize their information potential and drive top business imperatives. More than 5,800 enterprises depend on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

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.

## Unmatched 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 user interface, 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.



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

© 2016 Informatica LLC. All rights reserved. Informatica® and Put potential to work™ are trademarks or registered trademarks of Informatica in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks.