Informatica Data Integration Hub

Accelerate Transformation Projects with a Modern Hybrid Data Integration Hub.

Organizations across the globe are planning transformations to take advantage of new opportunities and counter new threats. In order to transform, they need to modernize their data infrastructure. Organizations are struggling to integrate their cloud, big data, and new analytics and applications in an organized way that enables them to deliver fresh, clean data for analytics and operational applications as well as to power new business initiatives.

Leveraging the cloud, big data, and innovations in analytics, databases, and applications with existing systems in an agile, efficient, and governed fashion requires a more intelligent solution to data integration than a traditional fragmented point-to-point ETL approach.

Modern data hub architecture for governance and sharing

Informatica Data Integration Hub empowers large organizations to embrace change and opportunity with a modern hybrid architecture for data integration that can connect all data and organize complex environments. It uses a self-service accessible publish/subscribe hub architecture optimized for any size and type of data and complex requirements for hybrid orchestration across systems in the cloud and on-premises. Data Integration Hub provides high productivity and efficiency with intelligent automation without compromising governance.

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. Support for mixed latencies means that each system gets the same fresh data exactly when it’s needed, whether in near real-time or at a set schedule. 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 simplify the creation of complex hybrid data solutions by orchestrating data integration of multiple clouds, big data, data quality, master data management, data as a service, and masking into consistent data publications for all analytics and operational systems.
Self-service empowerment for distributed teams

Enabling analysts and other less technical users to access and share data without requiring waiting for IT developer time dramatically improves productivity and fosters more positive business-IT collaboration. Data Integration Hub has simple step-by-step wizards and a web-based user interface that let business users set up their own schedules. Users can also set up data publications to be made available to other users. 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.

Pub/sub data hub to better govern data management

Using Informatica’s data hub approach, data sources are decoupled from destinations and respective data formats through a canonical topic model, abstracting complexity and providing a central place to go to understand what is happening with the entire system. The Hub Overview provides a visual tool to understand the relationships among the applications publishing data and subscribers to data topics. This easy-to-visualize lineage increases confidence in data and its quality. The ability to orchestrate data processing execution on Informatica Cloud, PowerCenter, and Hadoop together through a single system provides better visibility, monitoring, and management.

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 and multiple clouds

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 levels of productivity and manageability. Deployment of Data Integration Hub on AWS and Microsoft Azure is also supported to run Data Integration Hub in the cloud.

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 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 for defined retention periods. 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. 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, redundant Cloud data transfer, and API call charges, as well as virtually eliminating process dependencies, enabling downstream applications to readily meet the 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

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 among 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 governed data management through the hub and available Informatica Data Quality, all data can be certified and cleansed prior to publication. This virtually eliminates the risk of fragmentation or inconsistent 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 and canonical topic data models. By decoupling data sources and destinations, applications are less process-interdependent, enabling downstream applications to better meet the data delivery needs of business user.

**Wizards to enable self-service**

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 combined with self-service 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 100 percent focused on data because the world runs on data. Organizations need business solutions around data for the cloud, big data, real-time and streaming. Informatica is the world’s No. 1 provider of data management solutions, in the cloud, on-premise or in a hybrid environment. More than 7,000 organizations around the world turn to Informatica for data solutions that power their businesses.

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 about 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.

The Hub Overview provides a visual tool to understand data relationships and an overview of data lineage. This make it easy to comprehend the relationships between publications, topics, subscriptions and the connected applications.