

# Informatica Big Data Management

## Benefits

- Gain an intelligent approach to big data integration
- Integrate more data from more sources
- Accelerate developer productivity
- Get faster, flexible, repeatable execution of data pipelines on Apache Hadoop

## Ingest and Integrate More Big Data Quickly and Easily with Machine Learning

As the pace of business increases and organizations face overwhelming competitive pressure to transform their businesses, there is an opportunity to modernize and optimize data architectures to enable data to become a strategic asset for organizational decision-making. Apache Hadoop now gives organizations the luxury of using more data for analysis than ever before. But organizations are unable to turn most of their big data into trusted business insights due to antiquated and manual approaches of hand coding and code generation that leave most big data siloed, inconsistent, and incomplete.

## Integrate More Data Faster, from More Data Sources

Informatica Big Data Management® delivers high-throughput data ingestion and data integration processing so business analysts can get the data they need quickly. Hundreds of prebuilt, high-performance connectors, data integration transformations, and parsers enable virtually any type of data to be quickly ingested and processed on big data infrastructure, such as Apache Hadoop, NoSQL, and MPP appliances. Beyond prebuilt components and automation, Informatica provides dynamic mappings, dynamic schema support, and parameterization for programmatic and templated automation of data integration processes.

With a smart performance optimizer that is compatible with multiple processing engines, including MapReduce, Spark, Apache Tez, and Informatica Blaze, Informatica Big Data Management delivers maximum developer productivity, operational reusability, and data integration performance that ultimately shortens the time to value for business needs. It provides the gold standard in big data integration solutions so you can turn more big data into business value quickly.

“Informatica helps Tinkoff deliver on business expectations no matter how fast data volumes grow, how complex the data model is, and whichever data sources need to be integrated. By harnessing big data across the bank, Tinkoff is now uniquely positioned to target, acquire, and retain more customers.”

— Sergey Sotnichenko,  
Head of Data Warehousing,  
Tinkoff Bank

## Key Features

### Visual Development Interface

The Hadoop ecosystem is rapidly changing with new innovations continuously emerging in the open source community. Informatica Big Data Management builds on top of the open source Hadoop framework and preserves all the transformation logic in your data pipelines. This means developers can design once, without any specialized knowledge of Hadoop concepts and languages, and easily deploy data pipelines without having to rebuild each time Hadoop changes. As a result, Hadoop innovations are implemented faster with less impact and risk to production systems.

### Near Universal Data Connectivity

Nearly any type of big transaction data, including RDBMS, OLTP, OLAP, ERP, CRM, mainframe, and cloud, or interaction data like social media data, log files, machine sensor data, Hadoop, NoSQL formats, documents, and email can be accessed right out of the box.

### High-Speed Data Integration on Hadoop and Spark

A comprehensive library of prebuilt data integration transformation capabilities that run natively on Hadoop or Spark ensure that data can be processed at any scale. With powerful transformations right out of the box, combined with the visual development interface, developers can spend more time on business logic and less time on writing code.

### Faster Mass Ingestion and Extraction

Informatica Big Data Management generates hundreds of run-time data flows based on just a handful of design patterns using mass ingestion and mapping templates. You can easily parameterize these data flows to handle dynamic schemas such as web and machine log files, which are common to big data projects. This means you can quickly build data flows that are easy to maintain and resilient to changing schemas.

### Intelligent Data Parsing on Hadoop

Informatica Big Data Management makes it easy to access and parse complex, multistructured, hierarchical, unstructured, and industry-standard data such as web logs, JSON, XML, and machine device data. Machine learning infers the structure of data and creates bespoke parsers for repeatable use. Additional prebuilt parsers for market data and industry standards such as SWIFT, ACORD, HL7, HIPAA, and EDI are also available.

### Smart Performance Optimizer

Informatica’s Smart Optimizer enables you to execute the best engine for the highest performance, scalability, and resource utilization without having to rebuild data pipelines as new technologies emerge. For example, you can run data integration transformations using the Informatica Blaze engine in YARN, Spark, Hive on Tez, or MapReduce. Informatica Blaze is Informatica’s data processing engine integrated with YARN to provide intelligent data pipelining, job partitioning, job recovery, and high-performance scaling.

## About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

## SQL to Mapping Conversion

Informatica Big Data Management can translate ANSI-compliant SQL scripts, Informatica PowerCenter® pre- or post-SQL, and SQL override queries into optimized Informatica big data mappings that execute on Hadoop, maximizing reuse, simplifying maintenance, and preserving end-to-end data lineage.

## Flexible Serverless Deployment

Deploy and manage distributed resources automatically both on-premises and off-premises on Amazon Web Services Elastic MapReduce and Microsoft Azure HDInsight.

## Key Benefits

### Ingest Any Data

With the increased pace of business, data management professionals struggle to manually manage a growing volume and variety of data. Only Informatica's extensive library of prebuilt connectors enables organizations to ingest nearly any data. Using Informatica Big Data Management, your data scientists and analysts can focus on new data insights—not data integration—that your company can use to develop innovative products and services.

### Process and Deliver Data Anywhere

Data management professionals face a proliferation of next-generation data infrastructures. Only Informatica's certification and support of multiple on-premises and off-premises Hadoop distributions enable organizations to process and deliver data anywhere.

### Execute Faster

With a growing volume and variety of data, manual processes cannot scale. Manual processes take too long to execute and inhibit long-term maintainability. Informatica Big Data Management provides optimized run-time processing and simplified monitoring across multiple engines for faster, more flexible, and repeatable development and processing.

For more information, visit the [Informatica Big Data Management product page](#).



**Worldwide Headquarters** 2100 Seaport Blvd., Redwood City, CA 94063, USA Phone: 650.385.5000, Toll-free in the US: 1.800.653.3871

IN06\_0418\_02194

© Copyright Informatica LLC 2018. Informatica, Big Data Management, PowerCenter, and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners. The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.