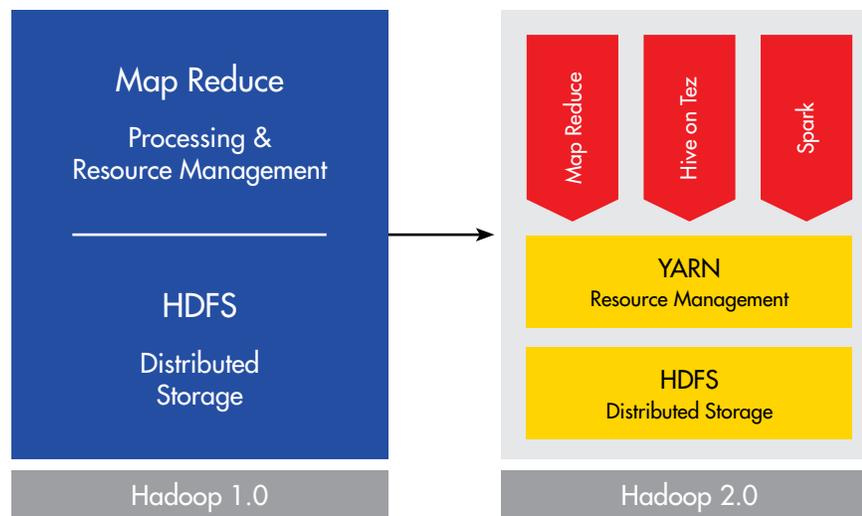


# Blaze Data Management Engine

## A Brief History of Hadoop

A Hadoop ecosystem is a platform for collecting, organizing, storing, and processing big data. In the early days of big data processing, Hadoop 1.0 allowed for storing big data in the Hadoop Distributed File System (HDFS) while using MapReduce as the programming framework for distributed processing of the data set and providing resource management capabilities. To process data, developers had to write MapReduce programs using Java or create Hive or Pig scripts which translated into a series of MapReduce jobs.



The constraints of a consolidated resource management and programming framework in Hadoop 1.0 led to the development of YARN (Yet Another Resource Negotiator). The next generation of Hadoop 2.0, YARN is a resource manager that separates the processing engine and resource management capabilities of MapReduce. YARN, often referred to as the data operating system of Hadoop, is designed to allow multiple, complex distributed applications to run on a multi-tenant platform, managing and monitoring workloads, implementing security controls and managing high availability features.

Informatica’s Big Data Management solution lets data engineers rapidly adopt new Hadoop innovations, such as Tez or Spark, without changes to existing data flows by uniquely providing optimized support for the best Hadoop processing engines for a broad spectrum of data management use cases.

Informatica leverages the MapReduce processing framework for non-compute intensive ETL/ELT batch processing. With Apache Tez, Informatica leverages the performance and scalability benefits for near real-time micro-batch processing workloads.

By leveraging Apache Spark, Informatica offers a publish/subscribe interaction pattern supporting a fully automated data delivery to both near real-time consumers as well as scheduled or event invoked delivery. Apache Spark is also integrated with Informatica Live Data Map, which continuously discovers, learns, and provides insight into all data assets, delivering an end-to-end holistic data governance solution.

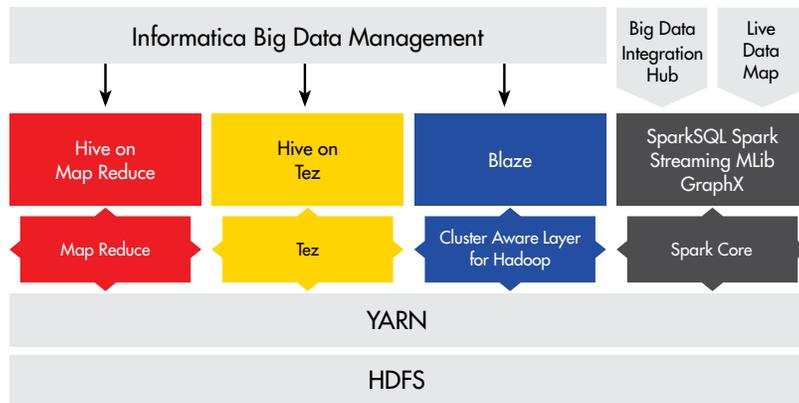
# Informatica Blaze

## 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.

Blaze is the industry's unique data processing engine integrated with YARN to provide intelligent data pipelining, job partitioning, job recovery, and scaling, which is optimized to deliver high performance data processing leveraging Informatica's propriety cluster aware data integration technology.

YARN, which provides the capability to build custom application frameworks on top of Hadoop to support multiple processing models, allowed the integration of Informatica's data transformation engine natively with Hadoop. Informatica Blaze is built using a memory-based data exchange framework which runs natively on YARN without the dependence of MapReduce or Hive and heeds the functional gaps of MapReduce. Informatica Blaze extends data processing capabilities on Hadoop by complementing Informatica's Big Data Management solutions supports multiple processing paradigms, such as MapReduce, Hive on Tez, Informatica Blaze, and Spark to execute each workload on the best possible processing engine.



## Key Benefits

- The Blaze distributed processing engine has the ability to scale and perform high speed data processing of large complex batch workloads via a natively embedded Informatica data transformation engine on Hadoop.
- The Cluster Aware Optimizer automatically determines the optimal batch processing layer at runtime across a variety of execution engines such as Informatica Blaze, Hive on Tez, and Hive on MapReduce using Hadoop or native Informatica server.

By combining the best of open source technology, YARN and 20 years of data management experience, the introduction of Informatica Blaze adds to Informatica's optimized support for multiple data processing frameworks by delivering flexible, scalable high performance data processing on Hadoop. This support ultimately provides organizations an end-to-end platform for optimized big data integration.

Informatica provides a comprehensive Big Data Management platform that delivers dynamic and optimized big data integration, end-to-end big data governance, and risk-centric data security. Big data integration delivers high-throughput data ingestion and at-scale processing so business analysts can make better decisions using next-generation analytics tools.



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)

© 2015 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.