

Data Warehouse Optimization with MapR and Informatica



Solution Overview

Organizations seek more and larger data sets in their data warehouses (DW) to extract more value. They derive better insights when analyzing a complete picture of enterprise-wide data. However, traditional data warehouses have a hard time keeping up with the growing data volumes. They do not cost-effectively scale to the levels organizations require, nor do they cost-effectively load newer forms of real-time data.

As a result, organizations make trade-offs like analyzing fewer data points via summaries or samples. In many cases, data older than a few months are discarded. This limited view of data inhibits the ability to gain important insights.

Optimizing the Data Warehouse

The MapR/Informatica data warehouse optimization (DWO) solution lets organizations cost-effectively add more data, more types of data, and more capabilities to their data warehouse environments. They can keep larger amounts of data for longer periods of time to gain deeper insights. Companies gain value from the many new, disparate data formats that are now available—cloud/mobile app data, social media, machine data, and many more. Organizations gain capabilities to keep up with higher speeds of incoming data, enabling real-time analytics for faster responsiveness to new insights.

The MapR/Informatica DWO solution adds efficiency to their data warehouse environment. Costly data warehouse upgrades are avoided by using the existing data warehouse for the highest-value analytical workloads. Portions of the data and workload can be offloaded to lower-cost systems running on commodity hardware. Tasks such as extract-transform-load (ETL), extract-load-transform (ELT), and data cleansing can be run on the lower-cost systems. Upon completion, the processed data can be reloaded into the data warehouse or delivered to data marts. Data can also be permanently retained in long-tail storage for performing large-scale analytics. By moving workloads off the data warehouse, the performance of high-value analytics is improved. Growth of the offloaded data is handled with linear, incremental scaling of a cluster of commodity servers.

The MapR/Informatica Enterprise Data Hub

The architecture of the MapR/Informatica DWO solution starts with the enterprise data hub (EDH), also referred to as a data lake. Analytics and storage are handled by the MapR Distribution for Apache™ Hadoop®. The data import, export, and processing are handled by Informatica running natively on Hadoop.

Solution Highlights

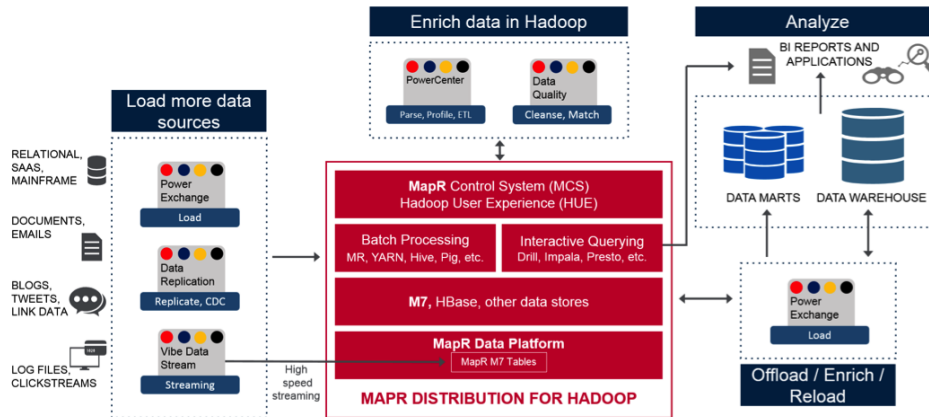
MapR and Informatica offer a data warehouse optimization solution that lets customers analyze larger volumes of data to get a complete picture of enterprise-wide data. The MapR and Informatica solution offers the features required by today's business-critical data warehouse environments.

Key Features

- The highest performance distributed computing platform for analytics and storage of virtually all data types
 - Hadoop for storing complex and unstructured data formats
 - NoSQL for storing non-relational, structured data formats
- Horizontal, incremental, linear scaling
- Easy-to-use interfaces for data integration, data quality, and other important data processing operations
- Business continuity features including high availability, disaster recovery, instant recovery, consistent point-in-time snapshots



Data Warehouse Optimization with MapR and Informatica



The MapR/Informatica EDH expands and complements an existing data warehouse environment, and reduces the load on data warehouses that lead to slower performance and higher costs. It enables storage and processing of data at any scale. The EDH supplements data warehouses with analytical tasks that focus on a variety of business needs. These include improvement of customer retention and acquisition, increasing operational efficiencies, enabling better products and service delivery, and generating new business insights.

To reduce the optimization effort, the MapR/Informatica DWO solution leverages existing and widely-available skillsets to ingest, process, and export data on Hadoop. And its low administrative overhead ensures IT teams can continue to meet stringent service level agreements (SLAs).

About Informatica

Informatica Corporation (Nasdaq: INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica to realize their information potential and drive top business imperatives. Informatica Vibe, the industry's first and only embeddable virtual data machine (VDM), powers the unique "Map Once. Deploy Anywhere." capabilities of the Informatica Platform. Worldwide, over 5,000 enterprises depend on Informatica to fully leverage their information assets from devices to mobile to social to Big Data residing on-premises, in the Cloud and across social networks.

About MapR

MapR delivers on the promise of Hadoop with a proven, enterprise-grade platform that supports a broad set of mission-critical and real-time production uses. MapR brings unprecedented dependability, ease-of-use and world-record speed to Hadoop, NoSQL, database and streaming applications in one unified big data platform.

Key Benefits

MapR Technologies and Informatica offer important benefits with their joint DWO solution. For storage and analysis, MapR provides the most advanced distribution for Hadoop. Informatica provides several powerful technologies for data integration, data quality, replication, streaming, and more.

Future-proofed data warehouse environment

- Derive more valuable insights with larger volume and higher velocity data
- Make data more valuable with a wide variety of data processing techniques
- Leverage innovations from a broad community

Increased productivity

- Reduce latency of integrating, processing, and analyzing big data
- Boost developer productivity for data processing by a factor of up to 5 times with Informatica
- Reuse existing analytical applications on the full read/write platform of MapR

Proven production-readiness

- Trust the extensive experience with production systems
- Ensure stringent SLAs are met
- Maintain business continuity to avoid costs of downtime

