Easing the transition to Hadoop

CISCO

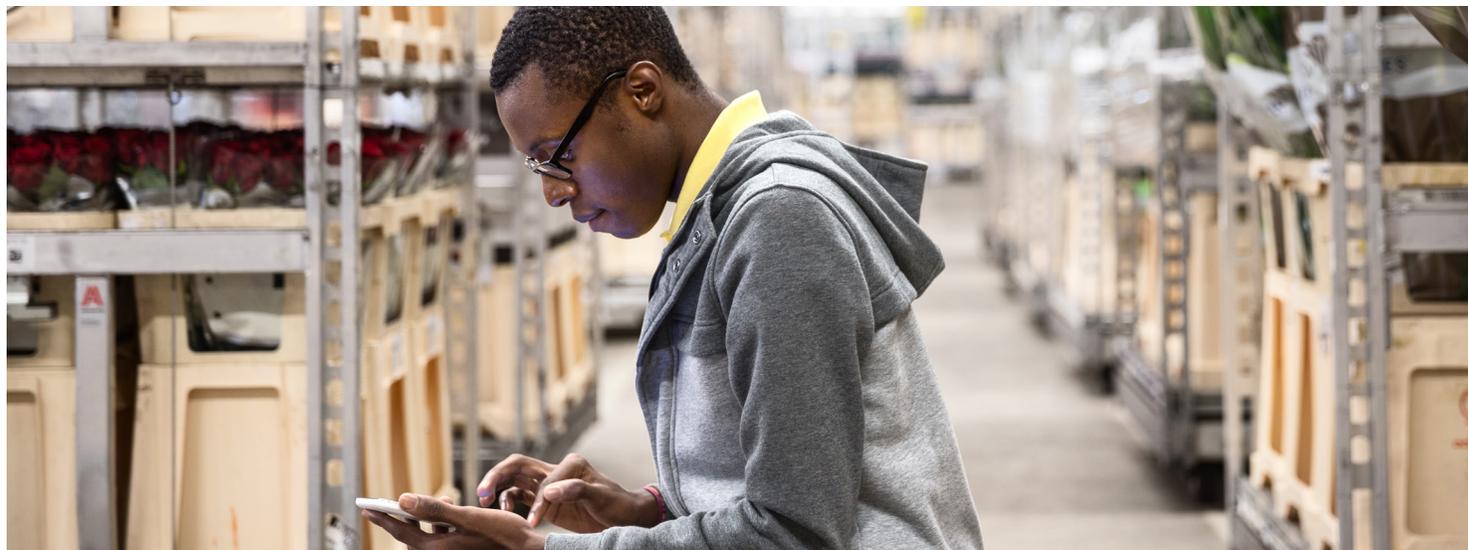**intel inside™ Xeon®**

Cisco and Intel® partnering in innovation

Virtualization and automation software can greatly reduce the time, complexity, and risk of offloading processing and storage from data warehouses to Hadoop.

Many companies are in the process of evaluating Hadoop—not just for big data analytics, but for a more cost-effective and scalable storage option than traditional data warehouses. But according to industry experts, the transition isn't always easy.

"People underestimate the effort and complexity of porting data to Hadoop," says Murthy Mathiprakasam, principal product marketing manager for Informatica, a leading provider of data integration software. "Hadoop is so new, not everyone understands how to use it effectively. And many try to move data manually, writing their own scripts and schema, which is time consuming and can be error prone."

To port data into Hadoop, it must first be collected and processed. And even after it has been moved, the data still requires a level of refinement. Get one of these steps wrong, and there can be undesired ripple effects. Bugs can emerge. There can be inconsistencies with the data. Or worse yet, business decisions can be made based on incomplete information and inaccurate assumptions.

FACILITATING GRADUAL TRANSITIONS

Fortunately, data virtualization and automation software can help ease Hadoop transitions.

"These things don't happen overnight," Mathiprakasam explains. "Automation software reduces the time and risk of porting the data from one place to the next. And virtualization enables the data to remain accessible during the transition, no matter where it is or where it will reside."

He points to the combination of Cisco® Data Virtualization, Informatica Big Data Edition, and the Intel® Xeon® processor-based Cisco Unified Computing System™ (Cisco UCS®) as an ideal platform for offloading processing and storage from data warehouses to Hadoop. **Informatica Big Data Edition** enables companies to run data transformation and data quality processes using a simple, visual development environment on Hadoop clusters installed on Cisco UCS servers. And the distributed data environments can be federated using Cisco Data Virtualization to provide business intelligence and analytics with a single point of access to all data.

"Hadoop is efficient and it's inexpensive," says Mathiprakasam. "But it isn't easy or simple. It's complex. The combination of Cisco Data Virtualization, Informatica software, and Cisco UCS can greatly reduce the time, effort, and risk of putting your data into a Hadoop cluster."

LEARN MORE

Informatica helps you make data ready for use in any way possible, so you can put truly great data at the center of everything you do. To register for Informatica events, webinars, white papers, and more, visit **dwoptimization.me**.

Informatica