Keeping Up With Big Data Innovation … Without Disrupting Your Business

In response to the challenges Big Data puts on traditional architectures, many organizations are turning to Apache Hadoop—whether they're evaluating its ability to increase efficiency and cut costs for storing and processing expensive workloads, standing up a pilot cluster to prove its value, or are already up and running in production.

Hadoop delivers great benefits to organizations seeking bigger insights and lower costs—and the ecosystem of related tools and solutions is constantly expanding to keep up as demands (real-time, self-service, etc.) and data growth (more sources, larger volumes) increase exponentially.

THE CHALLENGE OF RAPID INNOVATION

A rapidly-evolving ecosystem results in more choice and innovation, but added complexity, uncertainty and risk can get in the way of realizing the full potential of Big Data.

For example, Hadoop pioneers who originally chose MapReduce now face the daunting task of moving to new frameworks like Apache Spark. They—and everyone who comes after them—risk falling into the endless cycle of trying to keep up with bleeding edge technology to avoid obsolescence.

So, how do you pick the framework that is here to stay? You don't—because you can't. As the saying goes, the only thing constant in life is change—and even more so with Big Data.

Organizations want to take advantage of the powerful performance and scalability offered by rapidly evolving compute frameworks, but they also want to be confident that the applications they invest in today won't need to be rewritten, or require new skills, in 12 or 18 months.

"FUTURE-PROOFING" BECOMES ESSENTIAL

As a leader in enterprise software for more than 40 years—from Big Iron to Big Data, Syncsort knows a thing or two about adapting to change. That's why we designed Syncsort DMX-h, our industry-leading Big Data integration software, specifically with these challenges in mind. The foundation of the design is intelligent execution, which allows users to visually design data transformations once and deploy them anywhere—across Hadoop MapReduce, Apache Spark, Linux, UNIX, Windows—on premise or in the cloud—while achieving high levels of security, performance and scalability by taking advantage of native integration. As new frameworks inevitably emerge, DMX-h users will be able to deploy those same transformations there, too.

Because there are no required code changes when deploying on different frameworks, users can design sophisticated data transformations focused solely on business rules, without worrying about the underlying platform or execution framework—and without needing a new set of skills. This unique architecture insulates users from changes in the underlying technologies, "future-proofing" the process of collecting, blending, transforming, and distributing data. The user enjoys a consistent experience while still taking advantage of the powerful native performance of the evolving compute frameworks.

ADDING REAL-TIME TO THE MIX

The proliferation of tools and execution frameworks isn't the only change to manage. Data sources and types have exploded, as well. For example, we see increased demand for real-time analytics in industries such as financial services, healthcare, telecommunications and retail. The opportunities for these companies are significant. However, streaming telemetry data, real-time data from sensors and Internet of Things (IoT) use cases currently require managing different components in the technology stack. Businesses can also combine these real-time data sources with batch data to drive additional insights. But this, too, adds complexity.

For organizations to adopt new technologies and realize faster time to value, they need an easier way to manage all their processes. Many are interested in using a single software environment for streaming and batch processing, while taking advantage of the emerging compute frameworks like Apache Spark for analytics and the speed and resiliency of Apache Kafka for low-latency, fault-tolerant services.

Syncsort DMX-h integrates with messaging frameworks including Kafka to bring streaming and batch together through a single user interface. By removing obstacles to building streaming analytics applications on Hadoop, our solutions help close the gap between batch and real-time workloads. Organizations will be able to move data at scale in real-time across the enterprise to keep business systems in sync and speed-up slow batch processes with streaming ETL.

FOCUS ON SIMPLICITY, FLEXIBILITY

While Big Data innovation focuses on performance and scalability, flexibility and simplicity might prove just as important for long-term success. A single software environment that handles a broad spectrum of uses cases from operational efficiency to transformative applications *—and insulates users from change*—will let you enjoy benefits of innovation without disrupting your business. ■

SYNCSORT www.syncsort.com