CLOUDERA | NVIDIA

ACHIEVE 20X ETL PERFORMANCE GAINS

Learn how customers dramatically accelerate data workflows

Massive amounts of data and thin resources are a recipe for missed opportunity. Slow time-to-market, missing critical insights, delaying strategic business improvements, or waiting too long to deploy AI model improvements are all critical business issues that can be resolved by acceleration.

Harnessing today's advanced processing technology to overcome data slowdowns can make a big difference. The same kind of GPU acceleration that brings realistic 3D games to life and makes the deep learning revolution possible can also turbocharge your data engineering and analytics workloads.

How Apache Spark makes data ready faster

Data professionals use Apache Spark in support of data science, advanced analytics, and database queries at scale. Apache Spark's DataFrame API and relational query capabilities support crucial data-understanding, cleaning, and federation tasks. These structured-data and ETL workloads are an essential prerequisite to more advanced machine learning techniques, but they can be time-consuming.

That's where acceleration comes in. Apache Spark evaluates structured queries by constructing, optimizing, and executing a query plan, which helps the execution engine evaluate a query more efficiently. The RAPIDS Accelerator for Apache Spark integrates with Spark's query planner via a plugin. The plugin rewrites DataFrame query plans to evaluate accelerable operations with implementations that execute on the GPU. Operations that cannot be accelerated will run on the CPU with Spark's built-in implementations. The RAPIDS Accelerator plugin can also transfer operations between host and device memory, so both kinds of operations can work together transparently.

The bottom line is queries execute faster, velocity increases, time to insight is accelerated and costs and complexity are reduced, all with minimal or no engineering effort. GPU acceleration enables new use cases that once were not thought as possible due to the compute constraints.

Cloudera and NVIDIA make it possible

Cloudera and NVIDIA have partnered to bring GPU acceleration for Apache Spark 3.0+ to enterprises everywhere by shipping and supporting the RAPIDS Accelerator for Apache Spark in CDP Private Cloud Base. It's about helping data-driven organizations leap ahead without the slowdowns from data analytics bottlenecks.

Keep reading about GPU acceleration

See how the RAPIDS Accelerator for Apache Spark can accelerate data engineering, data preparation, and on-line analytic processing for predictive use cases in this ebook.

THE POWER OF CLOUDERA AND NVIDIA

Cloudera Data Platform (CDP) supports accelerating Apache Spark 3.0+ data engineering and analytics workloads with NVIDIA GPUs. This capability is jointly supported by Cloudera and NVIDIA and is generally available in Cloudera Data Platform 7.1.6+ for on-premises customers.

HOW FAST? TRY 20X FASTER.

CDP on NVIDIA GPUs accelerates ETL (extract, transform, load) pipelines to deliver data **20X faster** compared to running Apache Spark on CPUs only.

GREATER SPEED, LOWER COSTS

NVIDIA-Certified systems have lowered TCO by 50% to equivalent CPU-based systems.