GPU AGGREGATED 100.00% OF 146,594,624 MORTGAGES

Mortgage risk analysis.

NVIDIA.

Accelerating Analytics

Boost performance and cost savings for leading Apache Spark platforms.

Introduction

All kinds of enterprises use Apache Spark for ingesting raw data into data lakes, analyzing business processes, loading data into data warehouses, and preprocessing data at the start of machine learning pipelines.

The NVIDIA RAPIDS™ Accelerator for Apache Spark speeds up Apache Spark 3 for batch data pipelines. This acceleration can help organizations in many ways, including by improving throughput, solving problems at scale without sampling, scaling capabilities without enlarging data center footprints, and enabling analysts to ask complex questions at interactive latencies.

Applications

The **RAPIDS Accelerator for Apache Spark** software uses NVIDIA GPUs to accelerate processing of end-to-end data science and analytics pipelines. The software accelerates Spark jobs via a plug-in that integrates with all major Spark platforms—no code changes required. Operations that cannot be accelerated will continue to run on the CPU with Spark's built-in implementations.

Use Cases

Apache Spark is used for exploratory data analysis, feature generation, reporting, analytics, operational metrics, and more. End users include business, operations, product managers, and data analysts. Some specific use cases are demand forecasting, fraud detection, and price optimization.

Accelerate, Reclaim, and Save

With the NVIDIA RAPIDS Accelerator for Apache Spark, data analytics workloads can be completed 6X faster. This translates into a 5X savings on infrastructure costs and 6X less power for the same amount of work. For a typical enterprise, this can mean almost $125 million in savings with nearly 10GWh less energy consumed compared to running without GPU acceleration. NVIDIA has also developed an analysis tool that can estimate savings for any Apache Spark workload by analyzing the log file and calculating the degree of acceleration for each operation. This enables customers to determine how much they can expect to benefit from GPU acceleration of their Apache Spark workloads.

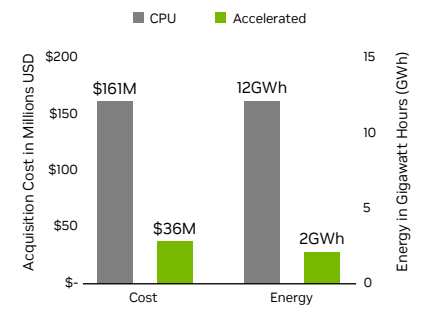To learn more, visit: **nvidia.com/sustainable-computing**
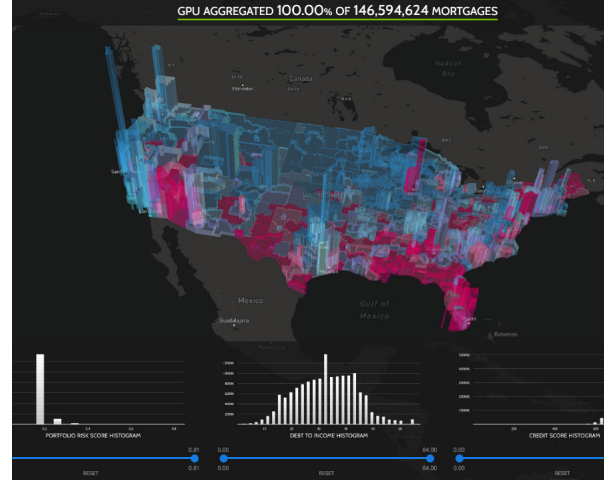
Key Points

> Apache Spark used by majority of Fortune 500

> Billions of CPU hours consumed annually

> 6X performance speedup

> $125 million and 10 gigawatt hours (GWh) saved annually

5X Lower Cost and 6X Less Energy

Based on NDS benchmark on Google Cloud Dataproc.

> "RAPIDS Accelerator and two NVIDIA A30 GPUs sustained the same production load for some workloads as a 200 CPU core test cluster in our Apache Spark workflow. When eventually deployed into production, we expect substantial energy efficiency gains by replacing hundreds of CPU cores with a minimal number of GPUs"

Eyal Hirsch,
Software Engineer, Taboola